

(19) World Intellectual Property
Organization
International Bureau



(43) International Publication Date
25 March 2004 (25.03.2004)

PCT

(10) International Publication Number
WO 2004/024084 A2

(51) International Patent Classification⁷: **A61K**

(21) International Application Number:
PCT/US2003/028626

MA 02166 (US). **CEOL, Craig** [—/US]; 43 Cohasset Street, Roslindale, MA 02131 (US). **ANDERSEN, Erik** [—/US]; 7 Asland Street, Apt.1, Somerville, MA 02144 (US).

(22) International Filing Date:
12 September 2003 (12.09.2003)

(74) Agent: **BIEKER-BRADY, Kristina, PH., D.**; Clark & Elbing LLP, 101 Federal Street, Boston, MA 02110 (US).

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/410,160 12 September 2002 (12.09.2002) US
60/437,821 2 January 2003 (02.01.2003) US

(81) Designated States (*national*): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

(63) Related by continuation (CON) or continuation-in-part (CIP) to earlier applications:
US 60/410,160 (CIP)
Filed on 12 September 2002 (12.09.2002)
US 60/437,821 (CIP)
Filed on 2 January 2003 (02.01.2003)

(84) Designated States (*regional*): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PT, RO, SE, SI, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

(71) Applicant (*for all designated States except US*): **MASSACHUSETTS INSTITUTE OF TECHNOLOGY** [US/US]; 77 Massachusetts Avenue, Cambridge, MA 02139 (US).

Published:

— *without international search report and to be republished upon receipt of that report*

(72) Inventors; and

(75) Inventors/Applicants (*for US only*): **HORVITZ, H., Robert, Ph., D.** [—/US]; 54 Maple Street, Auburndale,

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: RB PAHTWAY AND CHROMATIN REMODELING GENES THAT ANTAGONIZE *LET-60* RAS SIGNALING

(57) Abstract: In general, the invention provides methods and compositions useful in the treatment of a neoplasia. These compositions include new components of the Rb pathway that function in chromatin remodeling and antagonize Ras signaling.

WO 2004/024084 A2

5

10

25

30

characterized. Loss-of-function mutations in two functionally redundant pathways that are encoded by the class A and class B synthetic multivulva (synMuv) genes also cause a Muv phenotype.

In addition to LIN-35 Rb, other proteins with class B synMuv activity
5 are homologous to mammalian Rb-associated proteins. These other proteins include DPL-1 and EFL-1, homologs of DP and E2F transcription factors, LIN-53, a homolog of the Rb-binding proteins RbAp46 and RbAp48, HDA-1, a histone deacetylase homolog and HPL-2, a heterochromatin protein 1 homolog. The class B synMuv proteins act together to negatively regulate the
10 transcription of genes that promote vulval development. Initially, DPL-1 and EFL-1 heterodimers bind DNA at specific regulatory sequences of vulval cell-fate determination genes. DNA-bound DPL-1 and EFL-1 heterodimers recruit LIN-35 Rb, which in turn recruits proteins that act to remodel chromatin. One of these proteins, HDA-1, is predicted to deacetylate lysines of nucleosomal
15 histones. Deacetylation of lysine residues is required for their subsequent methylation. HPL-2, another protein that may be recruited by LIN-35 Rb, is expected to act like other HP1 family proteins and bind, via its chromodomain, to methylated lysine residues of nucleosomal histones.

Given the similarities that exist between *C. elegans* and mammalian Rb
20 and Ras pathways, *C. elegans* provides an efficient, inexpensive, and facile screening tool to identify novel clinical targets and chemotherapeutics useful in the treatment of neoplasia.

Summary of the Invention

25 The invention provides compositions useful in treating a neoplasia and methods for identifying chemotherapeutic agents.

In one aspect, the invention features a method for identifying a compound that treats a neoplasia, the method involves (a) contacting a cell containing a mutation in a Class B synMuv gene selected from the group
30 consisting of: *mep-1*, *lin(n3628)*, *lin(n4256)*, and *lin-65* and a second mutation

in a synthetic multivulval gene, or an ortholog thereof, with a candidate compound; and (b) detecting a phenotypic alteration in the contacted cell relative to a control cell; where a candidate compound that alters the phenotype of the contacted cell relative to the control cell is a compound that treats a neoplasia. In one embodiment, the cell is in a nematode. In another embodiment, the phenotypic alteration is an alteration in a multivulval phenotype. In another embodiment, the phenotypic alteration is an alteration in sterility. In another embodiment, the second mutation is in a synMuv class A gene. In another embodiment, the cell is an isolated mammalian cell. In another embodiment, the phenotypic alteration is a decrease in cell proliferation.

In another aspect, the invention provides a method for identifying a candidate compound that treats a neoplasia, the method involves (a) providing a cell having a mutation in a Class B synMuv gene selected from the group consisting of *mep-1*, *lin(n3628)*, *lin(n4256)*, and *lin-65* and having a second mutation in a synMuv nucleic acid or ortholog thereof; (b) contacting the cell with a candidate compound; and (c) detecting a decrease in proliferation of the cell contacted with the candidate compound relative to a control cell not contacted with the candidate compound, where a decrease in proliferation identifies the candidate compound as a candidate compound that treats a neoplasia. In one embodiment, the cell is in a nematode. In another embodiment, the decrease in proliferation is detected by detecting inhibition of a Muv phenotype. In another embodiment, the cell has a mutation in Dp, E2F, or histone deacetylase. In another embodiment, the cell is an isolated mammalian cell.

In another aspect, the invention provides a method of identifying a compound that treats a neoplasia, the method involves (a) providing a cell expressing a nucleic acid having at least 95% identity to a Class B synMuv gene selected from the group consisting of: *mep-1*, *lin(n3628)*, *lin(n4256)*, and *lin-65*; (b) contacting the cell with a candidate compound; and (c) monitoring

the expression of the nucleic acid, an alteration in the level of expression of the nucleic acid indicates that the candidate compound is a compound that treats a neoplasia. In one embodiment, the gene contains a reporter gene (e.g., *lacZ*, *gfp*, CAT, or luciferase). In another embodiment, expression is monitored by
5 assaying protein level. In another embodiment, the expression is monitored by assaying nucleic acid level. In yet another embodiment, the cell is in a nematode.

In another aspect, the invention features a method for identifying a candidate compound that treats a neoplasia, the method involves (a) providing
10 a cell expressing a Class B synMuv gene selected from the group consisting of: *mep-1*, *lin(n3628)*, *lin(n4256)*, and *lin-65*; (b) contacting the cell with a candidate compound; and (c) comparing the expression of the polypeptide in the cell contacted with the candidate compound to a control cell not contacted with the candidate compound, where an increase in the expression of the
15 polypeptide identifies the candidate compound as a candidate compound that treats a neoplasia. In one embodiment, the cell is in a nematode. In another embodiment, the expression is monitored with an immunological assay.

In another aspect, the invention features a method for identifying a candidate compound that treats a neoplasia, the method involves (a) providing
20 a cell expressing a Class B synMuv polypeptide selected from the group consisting of: MEP-1, LIN(n3628), LIN(n4256), and LIN-65, the method involves; (b) contacting the cell with a candidate compound; and (c) comparing the biological activity of the polypeptide in the cell contacted with the candidate compound to a control cell not contacted with the candidate
25 compound, where an increase in the biological activity of the polypeptide identifies the candidate compound as a candidate compound that treats a neoplasia. In another embodiment, the biological activity is monitored with an enzymatic assay. In another embodiment, the biological activity is monitored with an immunological assay. In yet another embodiment, the biological
30 activity is monitored with a nematode bioassay.

In another aspect, the invention features a method of identifying a nucleic acid target of class B synMuv biological activity, the method involves (a) mutagenizing a *C. elegans* containing mutations in a Class B synMuv gene selected from the group consisting of: *mep-1*, *lin(n3628)*, *lin(n4256)*, and *lin-65* and in a Class A synMuv gene; (b) allowing the *C. elegans* to reproduce; and (c) selecting a *C. elegans* containing a mutation that suppresses a synMuv phenotype; where the mutation identifies a nucleic acid target of class B synMuv biological activity.

In another aspect, the invention features a method of identifying a nucleic acid target of class B synMuv biological activity, the method involves (a) providing a microarray containing fragments of nematode nucleic acids; (b) contacting the microarray with detectably labeled nucleic acids derived from a nematode containing a mutation in a Class B synMuv gene selected from the group consisting of: *mep-1*, *lin(n3628)*, *lin(n4256)*, and *lin-65* gene; (c) detecting an alteration in the expression of at least one nucleic acid of a *C. elegans* containing a mutation in the Class B synMuv gene relative to the expression of the nucleic acid in a control nematode, where an alteration in the expression identifies the nucleic acid as a nucleic acid target of class B synMuv biological activity. In one embodiment, the *C. elegans* further contains a mutation in a second synMuv gene. In another embodiment, the *C. elegans* further contains a mutation in a gene that results in a Vulvaless (Vul) phenotype.

In another aspect, the invention features a method for identifying a nucleic acid that binds a synMuv class B polypeptide, the method involves (a) providing nucleic acids derived from a nematode cell; (b) crosslinking the nucleic acids and their associated proteins to form a nucleic acid-protein complex; (c) contacting the nucleic acid-protein complex with an antibody against a polypeptide selected from the group consisting of MEP-1, LIN(n3628), LIN(n4256), and LIN-65; (d) purifying the nucleic acid-protein complex using an immunological method; and (e) isolating the nucleic acid,

where the isolated nucleic acid is a nucleic acid that binds a synMuv class B polypeptide. In one embodiment, the method further involves the following steps: (f) detectably labeling the nucleic acid of step (e); (g) contacting a microarray containing *C. elegans* nucleic acid fragments with the detectably
5 labeled nucleic acid; and (h) detecting binding of the detectably labeled nucleic acid, where the binding identifies the nucleic acid as a nucleic acid that binds a synMuv class B polypeptide.

In another aspect, the invention provides a vector containing a nucleic acid having at least 95% identity to a Class B synMuv gene selected from the
10 group consisting of: *mep-1*, *lin(n3628)*, *lin(n4256)*, and *lin-65*. In one embodiment, the synMuv gene is *mep-1* (SEQ ID NO:2). In one embodiment, the synMuv gene contains a mutation selected from the group consisting of *n3680*, *n3702*, and *n3703*. In other embodiments, the synMuv gene is *lin(n3628)* (SEQ ID NO:24), *lin(n4256)* (SEQ ID NO:26), or *lin-65* (SEQ ID
15 NO:28).

In another aspect, the invention provides an isolated cell containing the vector of the previous aspect.

In a related aspect, the invention provides a nematode containing the nucleic acid of the previous aspect.

20 In another aspect, the invention provides a nematode containing a mutation in a Class B synMuv gene selected from the group consisting of: *mep-1*, *lin(n3628)*, *lin(n4256)*, and *lin-65*. In one embodiment, the mutation is a *mep-1* mutation selected from the group consisting of *n3680*, *n3702*, and *n3703*.

25 In another aspect, the invention features a purified nucleic acid containing a sequence that hybridizes under high stringency conditions to a Class B synMuv nucleic acid selected from the group consisting of: *mep-1*, *lin(n3628)*, *lin(n4256)*, and *lin-65*.

In another aspect, the invention features an antibody against a Class B synMuv polypeptide selected from the group consisting of: MEP-1, LIN(n3628), LIN(n4256), and LIN-65.

5 In another aspect, the invention provides a method for identifying a compound that treats a condition characterized by inappropriate cell death, the method involves (a) contacting a nematode containing a mutation in a Class B synMuv gene selected from the group consisting of: *mep-1*, *lin(n3628)*, *lin(n4256)*, and *lin-65* with a candidate compound; and (b) detecting a muv phenotype in the contacted nematode relative to a control nematode; where a
10 candidate compound that alters the phenotype of the contacted nematode relative to the control nematode is a compound that treats a condition characterized by inappropriate cell death. In one embodiment, the cell is in a nematode. In another embodiment, the alteration is an alteration in a synMuv phenotype.

15 In another aspect, the invention provides a method for identifying a compound that treats a neoplasia, the method involves (a) contacting a cell containing a mutation in a gene encoding KIAA1732 and a second mutation in a synMuv nucleic acid, or an ortholog thereof, with a candidate compound; (b) detecting a phenotypic alteration in the contacted cell relative to a control cell;
20 where a candidate compound that alters the phenotype of the contacted cell relative to the control cell is a compound that treats a neoplasia. In one embodiment, the synthetic multivulval gene is a synMuv class A gene. In another embodiment, the cell is an isolated mammalian cell. In another embodiment, the phenotypic alteration is a decrease in cell proliferation.

25 In another aspect, the invention features a method for identifying a candidate compound that treats a neoplasia, the method involves (a) providing a cell having a mutation in a nucleic acid encoding KIAA1732 and having a second mutation in a synMuv nucleic acid, or ortholog thereof; (b) contacting the cell with a candidate compound; and (c) detecting a decrease in
30 proliferation of the cell contacted with the candidate compound relative to a

control cell not contacted with the candidate compound, where a decrease in proliferation identifies the candidate compound as a candidate compound that treats a neoplasia. In one embodiment, the cell has a mutation in Dp, E2F, or histone deacetylase. In another embodiment, the cell is an isolated mammalian cell.

In another aspect, the invention provides a method of identifying a compound that treats a neoplasia, the method involves (a) providing a cell expressing a nucleic acid having at least 95% identity to a nucleic acid that encodes KIAA1732; (b) contacting the cell with a candidate compound; and (c) monitoring the expression of the nucleic acid, an alteration in the level of expression of the nucleic acid indicates that the candidate compound is a compound that treats a neoplasia. In one embodiment, the gene contains a reporter gene (e.g., *lacZ*, *gfp*, CAT, or luciferase). In another embodiment, expression is monitored by assaying protein level. In another embodiment, the expression is monitored by assaying nucleic acid level. In another embodiment, the cell is an isolated mammalian cell.

In another aspect, the invention provides a method for identifying a candidate compound that treats a neoplasia, the method involves (a) providing a cell expressing a KIAA1732 polypeptide; (b) contacting the cell with a candidate compound; and (c) comparing the expression of the polypeptide in the cell contacted with the candidate compound to a control cell not contacted with the candidate compound, where an increase in the expression of the polypeptide identifies the candidate compound as a candidate compound that treats a neoplasia. In one embodiment, the cell is an isolated mammalian cell. In another embodiment, the expression is monitored with an immunological assay.

In another aspect, the invention features a method for identifying a candidate compound that treats a neoplasia, the method involves (a) providing a cell expressing a KIAA1732 polypeptide; (b) contacting the cell with a candidate compound; and (c) comparing the biological activity of the

polypeptide in the cell contacted with the candidate compound to a control cell not contacted with the candidate compound, where an increase in the biological activity of the polypeptide identifies the candidate compound as a candidate compound that treats a neoplasia. In one embodiment, the biological activity is monitored with an enzymatic assay. In another embodiment, the biological activity is monitored with an immunological assay. In another embodiment, the biological activity is methyl transferase activity.

In another aspect, the invention features a method for identifying a nucleic acid that binds KIAA1732, the method involves (a) providing nucleic acids derived from a mammalian cell; (b) crosslinking the nucleic acids and their associated proteins to form a nucleic acid-protein complex; (c) contacting the nucleic acid-protein complex with an anti-KIAA1732 antibody; (d) purifying the nucleic acid-protein complex using an immunological method; and (e) isolating the nucleic acid, where the isolated nucleic acid is a nucleic acid that binds KIAA1732. In one embodiment, the method further involves the following steps: (f) detectably labeling the nucleic acid of step (e); (g) contacting a microarray containing human nucleic acid fragments with the detectably labeled nucleic acid; and (h) detecting binding of the detectably labeled nucleic acid, where the binding identifies the nucleic acid as a nucleic acid that binds KIAA1732.

In another aspect, the invention provides a vector containing a nucleic acid having at least 95% identity to SEQ ID NO:36.

In another aspect, the invention provides an isolated cell containing the vector of the previous aspect.

In another aspect, the invention provides a method for identifying a compound that treats a neoplasia, the method involves (a) contacting a nematode containing a mutation in a Class C synMuv gene selected from the group consisting of *trr-1*, *hat-1*, *epc-1*, and *ssl-1* with a candidate compound; and (b) detecting an altered phenotype in the contacted nematode relative to a control nematode; where a candidate compound that alters the phenotype of

the contacted nematode relative to the control nematode is a compound that treats a neoplasia. In one embodiment, the alteration is an alteration in vulval phenotype. In another embodiment, the alteration is an alteration in sterility. In another embodiment, the synMuv class C gene is *trr-1*. In another
5 embodiment, the mutations are selected from the group consisting of *n3630*, *n3637*, *n3704*, *n3708*, *n3709*, and *n3712*.

In another aspect, the invention provides a method for identifying a candidate compound that treats a neoplasia, the method involves (a) providing a cell having a mutation in a Class C synMuv gene selected from the group
10 consisting of *trr-1*, *hat-1*, *epc-1*, and *ssl-1* and having a second mutation in a synMuv nucleic acid or ortholog thereof; (b) contacting the cell with a candidate compound; and (c) detecting a decreased proliferation of the cell contacted with the candidate compound relative to a control cell not contacted with the candidate compound, where a decrease in proliferation identifies the
15 candidate compound as a candidate compound that treats a neoplasia. In one embodiment, the cell is in a nematode. In another embodiment, the nematode displays an alteration in a synMuv phenotype. In another embodiment, the cell contains a mutation in a class A or class B synMuv gene.

In another aspect, the invention provides a method for identifying a
20 compound that treats a neoplasia, the method involves (a) contacting a nematode containing a mutation in a Class C synMuv gene selected from the group consisting of *trr-1*, *hat-1*, *epc-1*, and *ssl-1* and a second mutation in a Class A synthetic multivulval gene with a candidate compound; and (b) detecting an altered phenotype in the contacted nematode relative to a control
25 nematode; where a candidate compound that alters the phenotype of the contacted nematode relative to the control nematode is a compound that treats a neoplasia. In one embodiment, the alteration is an alteration in synMuv phenotype. In another embodiment, the alteration is an alteration in sterility.

In another aspect, the invention provides a method for identifying a
30 compound that treats a neoplasia, the method involves (a) contacting a

nematode containing a mutation in a Class C synMuv gene selected from the group consisting of *trr-1*, *hat-1*, *epc-1*, and *ssl-1* and a second mutation in a Class B synthetic multivulval gene with a candidate compound; (b) detecting an altered phenotype in the contacted nematode relative to a control nematode; 5 where a candidate compound that alters the phenotype of the contacted nematode relative to the control nematode is a compound that treats a neoplasia. In another embodiment, the alteration is an alteration in synMuv phenotype. In another embodiment, the alteration is an alteration in sterility. In another aspect, the invention features a method for identifying a candidate 10 compound that treats a neoplasia, the method involves (a) providing a cell having a mutation in a Class C synMuv gene selected from the group consisting of *trr-1*, *hat-1*, *epc-1*, and *ssl-1* and having a second mutation in a synMuv gene or ortholog thereof; (b) contacting the cell with a candidate compound; and (c) detecting a decreased proliferation of the cell contacted 15 with the candidate compound relative to a control cell not contacted with the candidate compound, where a decrease in proliferation identifies the candidate compound as a candidate compound that treats a neoplasia. In one embodiment, the cell is in a nematode. In another embodiment, the nematode displays an alteration in a synMuv phenotype.

20 In another aspect, the invention provides a method of identifying a compound that treats a neoplasia, the method involves (a) providing a cell expressing a nucleic acid having at least 95% identity to a Class C synMuv nucleic acid selected from the group consisting of *trr-1*, *hat-1*, *epc-1*, and *ssl-1*; (b) contacting the cell with a candidate compound; and (c) monitoring the 25 expression of the nucleic acid, an alteration in the level of expression of the nucleic acid indicates that the candidate compound is a compound that treats a neoplasia. In one embodiment, the gene contains a reporter gene. In another embodiment, the reporter gene contains *lacZ*, *gfp*, *CAT*, or luciferase. In another embodiment, the expression is monitored by assaying protein level. In

yet another embodiment, the expression is monitored by assaying nucleic acid level. In yet another embodiment, the nucleic acid is in a nematode.

In another aspect, the invention provides a method for identifying a candidate compound that treats a neoplasia, the method involves (a) providing
5 a cell expressing a a Class C synMuv polypeptide selected from the group consisting of TRR-1, HAT-1, EPC-1, and SSL-1 polypeptide; (b) contacting the cell with a candidate compound; and (c) comparing the expression of the polypeptide in the cell contacted with the candidate compound to a control cell not contacted with the candidate compound, where an increase in the
10 expression of the polypeptide identifies the candidate compound as a candidate compound that treats a neoplasia. In one embodiment, the cell is in a nematode. In another embodiment, the expression is monitored with an immunological assay.

In another aspect, the invention provides a method for identifying a
15 candidate compound that treats a neoplasia, the method involves (a) providing a cell expressing a Class C synMuv polypeptide selected from the group consisting of TRR-1, HAT-1, EPC-1, and SSL-1; (b) contacting the cell with a candidate compound; and (c) comparing the biological activity of the polypeptide in the cell contacted with the candidate compound to a control cell
20 not contacted with the candidate compound, where an increase in the biological activity of the polypeptide identifies the candidate compound as a candidate compound that treats a neoplasia. In one embodiment, the cell is in a nematode. In another embodiment, the biological activity is monitored with an enzymatic assay. In another embodiment, the biological activity is monitored
25 with an immunological assay.

In another aspect, the invention provides a method of identifying a nucleic acid target of a synMuv Class C polypeptide, the method involves (a) mutagenizing a *C. elegans* containing a first mutation in a Class C synMuv gene selected from the group consisting of *trr-1*, *hat-1*, *epc-1*, and *ssl-1* and a
30 second mutation in a Class A or Class B synMuv gene; (b) allowing the *C.*

elegans to reproduce; (c) selecting a *C. elegans* containing a mutation that suppresses a synMuv phenotype; where the mutation identifies a nucleic acid target of a synMuv class C polypeptide. In one embodiment, the second mutation is in a class A synMuv gene. In another embodiment, the second
5 mutation is in a Class B synMuv gene.

In another aspect, the invention provides a method for identifying a nucleic acid target of a synMuv Class C polypeptide, the method involves (a) providing a *C. elegans* containing a mutations in a Class C synMuv gene selected from the group consisting of *trr-1*, *hat-1*, *epc-1*, and *ssl-1*; (b) growing
10 the *C. elegans* on bacteria expressing a dsRNA; and (c) identifying a dsRNA that suppresses a synMuv phenotype; where the dsRNA identifies a nucleic acid target of a synMuv class C polypeptide.

In another aspect, the invention provides a method for identifying a nucleic acid target of a synMuv class C polypeptide, the method involves (a)
15 providing a *C. elegans* containing mutations in a Class C synMuv gene selected from the group consisting of *trr-1*, *hat-1*, *epc-1*, and *ssl-1* and in a Class A or Class B synMuv gene; (b) growing the *C. elegans* on bacteria expressing a dsRNA; and (c) identifying a dsRNA that suppresses a synMuv phenotype; where the dsRNA identifies a nucleic acid target of a synMuv class C
20 polypeptide.

In another aspect, the invention features a method of identifying a nucleic acid whose expression is modulated by a synMuv class C polypeptide, the method involves (a) providing a microarray containing fragments of nematode nucleic acids; (b) contacting the microarray with detectably labeled
25 nucleic acids derived from a nematode containing a mutation in a Class C synMuv gene selected from the group consisting of *trr-1*, *hat-1*, *epc-1*, and *ssl-1* gene; (c) detecting an alteration in the expression of at least one nucleic acid of a *C. elegans* containing a mutation in the synMuv class C gene relative to the expression of the nucleic acid in a control nematode, where an alteration in
30 the expression identifies the nucleic acid as a nucleic acid modulated by a

synMuv class C polypeptide. In one embodiment, the *C. elegans* further contains a mutation in a synMuv A or synMuv B gene. In another embodiment, the *C. elegans* further contains a mutation in a gene that results in a Vulvaless (Vul) phenotype. In another embodiment, the gene encodes LET-60.

In another aspect, the invention provides a method for identifying a nucleic acid target of a synMuv class C polypeptide, the method involves (a) providing nucleic acids derived from a nematode cell; (b) crosslinking the nucleic acids and their associated proteins to form a nucleic acid-protein complex; (c) contacting the nucleic acid-protein complex with an antibody that binds a polypeptide selected from the group consisting of TRR-1, HAT-1, EPC-1, AND SSL-1; (d) purifying the nucleic acid-protein complex using an immunological method; and (e) isolating the nucleic acid, where the isolated nucleic acid is a nucleic acid that binds a synMuv class C polypeptide. In another embodiment, further containing the following steps: (f) detectably labeling the nucleic acid of step (e); (g) contacting the detectably labeled nucleic acid with a microarray containing *C. elegans* nucleic acid fragments; and (h) detecting binding of the detectably labeled nucleic acid, where the binding identifies the nucleic acid as a nucleic acid target of a synMuv class C polypeptide.

By "binds" is meant a compound or antibody which recognizes and binds a polypeptide of the invention, but which does not substantially recognize and bind other different molecules in a sample, for example, a biological sample, which naturally includes a polypeptide of the invention.

By "cell" is meant a single-cellular organism, cell from a multi-cellular organism, or it may be a cell contained in a multi-cellular organism.

By "derived from" is meant isolated from or having the sequence of a naturally-occurring sequence (e.g., a cDNA, genomic DNA, synthetic, or combination thereof).

“Differentially expressed” means a difference in the expression level of a nucleic acid. This difference may be either an increase or a decrease in expression, when compared to control conditions.

By “*epc-1* nucleic acid” is meant a synMuv Class C nucleic acid
5 substantially identical to Y111B2A.11, which is identified by *C. elegans* cosmid name and open reading frame number.

By “EPC-1 polypeptide” is meant an amino acid sequence substantially identical to a polypeptide expressed by an *epc-1* nucleic acid that that functions in vulval development and associates with a MYST family histone
10 acetyltransferase.

By “fragment” is meant a portion of a protein or nucleic acid that is substantially identical to a reference protein or nucleic acid (e.g., one of those listed in Tables 2 or 3), and retains at least 50% or 75%, more preferably 80%, 90%, or 95%, or even 99% of the biological activity of the reference protein or
15 nucleic acid using a nematode bioassay as described herein or a standard biochemical or enzymatic assay.

By “hybridize” is meant pair to form a double-stranded molecule between complementary polynucleotide sequences (e.g., genes listed in Tables 1-4 and 7), or portions thereof, under various conditions of stringency. (See,
20 e.g., Wahl, G. M. and S. L. Berger (1987) *Methods Enzymol.* 152:399; Kimmel, A. R. (1987) *Methods Enzymol.* 152:507) For example, stringent salt concentration will ordinarily be less than about 750 mM NaCl and 75 mM trisodium citrate, preferably less than about 500 mM NaCl and 50 mM trisodium citrate, and most preferably less than about 250 mM NaCl and 25
25 mM trisodium citrate. Low stringency hybridization can be obtained in the absence of organic solvent, e.g., formamide, while high stringency hybridization can be obtained in the presence of at least about 35% formamide, and most preferably at least about 50% formamide. Stringent temperature conditions will ordinarily include temperatures of at least about 30°C, more
30 preferably of at least about 37°C, and most preferably of at least about 42°C.

Varying additional parameters, such as hybridization time, the concentration of detergent, e.g., sodium dodecyl sulfate (SDS), and the inclusion or exclusion of carrier DNA, are well known to those skilled in the art. Various levels of stringency are accomplished by combining these various conditions as needed.

5 In a preferred embodiment, hybridization will occur at 30°C in 750 mM NaCl, 75 mM trisodium citrate, and 1% SDS. In a more preferred embodiment, hybridization will occur at 37°C in 500 mM NaCl, 50 mM trisodium citrate, 1% SDS, 35% formamide, and 100 µg/ml denatured salmon sperm DNA (ssDNA). In a most preferred embodiment, hybridization will occur at 42°C in
10 250 mM NaCl, 25 mM trisodium citrate, 1% SDS, 50% formamide, and 200 µg/ml ssDNA. Useful variations on these conditions will be readily apparent to those skilled in the art.

For most applications, washing steps that follow hybridization will also vary in stringency. Wash stringency conditions can be defined by salt
15 concentration and by temperature. As above, wash stringency can be increased by decreasing salt concentration or by increasing temperature. For example, stringent salt concentration for the wash steps will preferably be less than about 30 mM NaCl and 3 mM trisodium citrate, and most preferably less than about 15 mM NaCl and 1.5 mM trisodium citrate. Stringent temperature conditions
20 for the wash steps will ordinarily include a temperature of at least about 25°C, more preferably of at least about 42°C, and most preferably of at least about 68°C. In a preferred embodiment, wash steps will occur at 25°C in 30 mM NaCl, 3 mM trisodium citrate, and 0.1% SDS. In a more preferred
25 embodiment, wash steps will occur at 42°C in 15 mM NaCl, 1.5 mM trisodium citrate, and 0.1% SDS. In a most preferred embodiment, wash steps will occur at 68°C in 15 mM NaCl, 1.5 mM trisodium citrate, and 0.1% SDS. Additional variations on these conditions will be readily apparent to those skilled in the art. Hybridization techniques are well known to those skilled in the art and are described, for example, in Benton and Davis (*Science* 196:180, 1977);
30 Grunstein and Hogness (*Proc. Natl. Acad. Sci., USA* 72:3961, 1975); Ausubel

et al. (*Current Protocols in Molecular Biology*, Wiley Interscience, New York, 2001); Berger and Kimmel (*Guide to Molecular Cloning Techniques*, 1987, Academic Press, New York); and Sambrook et al., *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory Press, New York.

5 By "*hat-1* nucleic acid" is meant a a synMuv Class C nucleic acid substantially identical to VC5.4, which is identified by *C. elegans* cosmid name and open reading frame number.

By "HAT-1 polypeptide" is meant an amino acid sequence substantially identical to a polypeptide expressed by a *hat-1* nucleic acid that functions in
.. 10 . . vulval development and contains a chromodomain and an acetyltransferase catalytic domain.

By "*lin(n3628)* nucleic acid" is meant a nucleic acid substantially identical to SEQ ID NO:24 that encodes a histone methyltransferase.

By "LIN(n3628) polypeptide" is meant an amino acid sequence having
15 substantial identity to a polypeptide expressed by a *lin(n3628)* nucleic acid that has histone methyltransferase activity and includes a SET domain.

By "*lin(n4256)* nucleic acid" is meant a synMuv class B nucleic acid substantially identical to SEQ ID NO:27.

By "LIN(n4256) polypeptide" is meant an amino acid sequence having
20 substantial identity to a polypeptide expressed by a *lin(n4256)* nucleic acid and having histone methyltransferase activity.

By "*lin-65* nucleic acid" is meant a synMuv class B nucleic acid substantially identical to SEQ ID NO:28.

By "LIN-65 polypeptide" is meant an amino acid sequence having
25 substantial identity to a polypeptide expressed by a *lin-65* nucleic acid that is rich in acidic amino acids.

By "immunological assay" is meant an assay that relies on an immunological reaction, for example, antibody binding to an antigen. Examples of immunological assays include ELISAs, Western blots,
30 immunoprecipitations, and other assays known to the skilled artisan.

By "isolated polynucleotide" is meant a nucleic acid (e.g., a DNA) that is free of the genes which, in the naturally-occurring genome of the organism from which the nucleic acid molecule of the invention is derived, flank the gene. The term therefore includes, for example, a recombinant DNA that is
5 incorporated into a vector; into an autonomously replicating plasmid or virus; or into the genomic DNA of a prokaryote or eukaryote; or that exists as a separate molecule (for example, a cDNA or a genomic or cDNA fragment produced by PCR or restriction endonuclease digestion) independent of other sequences. In addition, the term includes an RNA molecule that is transcribed
10 from a DNA molecule, as well as a recombinant DNA that is part of a hybrid gene encoding additional polypeptide sequence.

By an "isolated polypeptide" is meant a polypeptide of the invention that has been separated from components that naturally accompany it. Typically, the polypeptide is isolated when it is at least 60%, by weight, free
15 from the proteins and naturally-occurring organic molecules with which it is naturally associated. Preferably, the preparation is at least 75%, more preferably at least 90%, and most preferably at least 99%, by weight, a polypeptide of the invention. An isolated polypeptide of the invention may be obtained, for example, by extraction from a natural source, by expression of a
20 recombinant nucleic acid encoding such a polypeptide; or by chemically synthesizing the protein. Purity can be measured by any appropriate method, for example, column chromatography, polyacrylamide gel electrophoresis, or by HPLC analysis.

By "KIAAA1732 nucleic acid" is meant a human nucleic acid sequence
25 having substantial identity to SEQ ID NO:30 and encoding a histone methyltransferase.

By "KIAAA1732 polypeptide" is meant an amino acid sequence encoded by a nucleic acid substantially identical to SEQ ID NO:30, having histone methyltransferase activity, and including a SET domain.

By “*mep-1* nucleic acid” is meant a a synMuv Class B nucleic acid substantially identical to M04B2.1, which is identified by *C. elegans* cosmid name and open reading frame number.

By “MEP-1 polypeptide” is meant an amino acid sequence substantially
 5 identical to a polypeptide expressed by a *mep-1* nucleic acid that functions in vulval development and contains multiple Zn finger motifs.

By “multivulva” is meant having one vulva and one additional vulva-like structure.

By “nucleic acid” is meant an oligomer or polymer of ribonucleic acid
 10 or deoxyribonucleic acid, or analog thereof. This term includes oligomers consisting of naturally occurring bases, sugars, and intersugar (backbone) linkages as well as oligomers having non-naturally occurring portions which function similarly. Such modified or substituted oligonucleotides are often preferred over native forms because of properties such as, for example,
 15 enhanced cellular uptake and increased stability in the presence of nucleases.

Specific examples of some preferred nucleic acids envisioned for this invention may contain phosphorothioates, phosphotriesters, methyl
 phosphonates, short chain alkyl or cycloalkyl intersugar linkages or short chain heteroatomic or heterocyclic intersugar linkages. Most preferred are those with
 20 $\text{CH}_2\text{—NH—O—CH}_2$, $\text{CH}_2\text{—N(CH}_3\text{)—O—CH}_2$, $\text{CH}_2\text{—O—N(CH}_3\text{)—CH}_2$, $\text{CH}_2\text{—N(CH}_3\text{)—N(CH}_3\text{)—CH}_2$ and $\text{O—N(CH}_3\text{)—CH}_2\text{—CH}_2$ backbones (where phosphodiester is O—P—O—CH_2). Also preferred are oligonucleotides having morpholino backbone structures (Summerton, J.E. and Weller, D.D., U.S. Pat. No: 5,034,506). In other preferred embodiments, such
 25 as the protein-nucleic acid (PNA) backbone, the phosphodiester backbone of the oligonucleotide may be replaced with a polyamide backbone, the bases being bound directly or indirectly to the aza nitrogen atoms of the polyamide backbone (P.E. Nielsen et al. *Science* 199: 254, 1997). Other preferred oligonucleotides may contain alkyl and halogen-substituted sugar moieties
 30 comprising one of the following at the 2' position: OH, SH, SCH_3 , F, OCN,

O(CH₂)_nNH₂ or O(CH₂)_nCH₃, where n is from 1 to about 10; C₁ to C₁₀ lower alkyl, substituted lower alkyl, alkaryl or aralkyl; Cl; Br; CN; CF₃; OCF₃; O-, S-, or N-alkyl; O-, S-, or N-alkenyl; SOCH₃; SO₂CH₃; ONO₂; NO₂; N₃; NH₂; heterocycloalkyl; heterocycloalkaryl; aminoalkylamino; polyalkylamino; substituted silyl; an RNA cleaving group; a conjugate; a reporter group; an intercalator; a group for improving the pharmacokinetic properties of an oligonucleotide; or a group for improving the pharmacodynamic properties of an oligonucleotide and other substituents having similar properties. Oligonucleotides may also have sugar mimetics such as cyclobutyls in place of the pentofuranosyl group.

Other preferred embodiments may include at least one modified base form. Some specific examples of such modified bases include 2-(amino)adenine, 2-(methylamino)adenine, 2-(imidazolylalkyl)adenine, 2-(aminoalkylamino)adenine, or other heterosubstituted alkyladenines.

By "ortholog" is meant a polypeptide or nucleic acid molecule of an organism that is highly related to a reference protein, or nucleic acid sequence, from another organism. An ortholog is functionally related to the reference protein or nucleic acid sequence. In other words, the ortholog and its reference molecule would be expected to fulfill similar, if not equivalent, functional roles in their respective organisms. It is not required that an ortholog, when aligned with a reference sequence, have a particular degree of amino acid sequence identity to the reference sequence. A protein ortholog might share significant amino acid sequence identity over the entire length of the protein, for example, or, alternatively, might share significant amino acid sequence identity over only a single functionally important domain of the protein. Such functionally important domains may be defined by genetic mutations or by structure-function assays. Orthologs may be identified using methods provided herein. The functional role of an ortholog may be assayed using methods well known to the skilled artisan, and described herein. For example, function might be assayed *in vivo* or *in vitro* using a biochemical, immunological, or enzymatic

assay; transformation rescue, or in a nematode bioassay for the effect of gene inactivation on nematode phenotype (e.g., fertility), as described herein.

Alternatively, bioassays may be carried out in tissue culture; function may also be assayed by gene inactivation (e.g., by RNAi, siRNA, or gene knockout), or
5 gene over-expression, as well as by other methods.

By “polypeptide” is meant any chain of amino acids, or analogs thereof, regardless of length or post-translational modification (for example, glycosylation or phosphorylation).

By “positioned for expression” is meant that the polynucleotide of the
10 invention (e.g., a DNA molecule) is positioned adjacent to a DNA sequence that directs transcription and translation of the sequence (i.e., facilitates the production of, for example, a recombinant polypeptide of the invention, or an RNA molecule).

By “purified antibody” is meant an antibody that is at least 60%, by
15 weight, free from proteins and naturally-occurring organic molecules with which it is naturally associated. Preferably, the preparation is at least 75%, more preferably 90%, and most preferably at least 99%, by weight, antibody. A purified antibody of the invention may be obtained, for example, by affinity chromatography using a recombinantly-produced polypeptide of the invention
20 and standard techniques.

By “specifically binds” is meant a compound or antibody that recognizes and binds a polypeptide of the invention, but which does not substantially recognize and bind other molecules in a sample, for example, a biological sample, which naturally includes a polypeptide of the invention.

By “*ssl-1* nucleic acid” is meant a nucleic acid substantially identical to
25 SEQ ID NO:21, which is identified by *C. elegans* cosmid name and open reading frame number.

By “SSL-1 polypeptide” is meant an amino acid sequence substantially identical to a polypeptide expressed by a *ssl-1* nucleic acid that functions in

embryonic development and has homology to p400 a SWI2/SNF2 family member having ATPase activity .

By “synthetic multivulva (synMuv) gene” is meant a gene that when mutated, interacts synergistically with a second synMuv gene to cause a synthetic multivulval phenotype. For example, *trr-1* and *mep-1* are synMuv genes because worms containing a mutation in *trr-1* or *mep-1*, and also having a mutation in *lin-15A* (e.g., *lin-15A(n767)*) display a synthetic multivulval phenotype.

By “*trr-1* nucleic acid” is meant a nucleic acid substantially identical to SEQ.ID.NO:12, which is identified by *C. elegans* cosmid name and open reading frame number. Nucleic acid and polypeptide sequence information is available at wormbase (www.wormbase.org), a central repository of data on *C. elegans*.

By “TRR-1 polypeptide” is meant an amino acid sequence substantially identical to a polypeptide expressed by a *trr-1* nucleic acid that functions in transcriptional regulation and vulval development.

“Therapeutic compound” means a substance that has the potential of affecting the function of an organism. Such a compound may be, for example, a naturally occurring, semi-synthetic, or synthetic agent. For example, the test compound may be a drug that targets a specific function of an organism. A test compound may also be an antibiotic or a nutrient. A therapeutic compound may decrease, suppress, attenuate, diminish, arrest, or stabilize the development or progression of disease, disorder, or infection in a eukaryotic host organism.

The invention provides a number of targets that are useful for the development of highly specific drugs to treat neoplasia or a disorder characterized by the misregulation of the cell cycle (e.g., a hyperproliferative disorder). In addition, the methods of the invention provide a facile means to identify therapies that are safe for use in eukaryotic host organisms (i.e., compounds that do not adversely affect the normal development, physiology,

or fertility of the organism). In addition, the methods of the invention provide a route for analyzing virtually any number of compounds for effects on cell proliferation and cell cycle regulation with inexpensively and with high-volume throughput in a living animal.

5 Other features and advantages of the invention will be apparent from the detailed description, and from the claims.

The invention provides methods and compositions useful in treating a neoplasia and in identifying chemotherapeutic agents. Other features and advantages of the invention will be apparent from the detailed description, and
10 from the claims.

Brief Description of the Drawings

Figure 1A is a schematic diagram the location of *mep-1* on the LGIV physical map in between *sem-3* and *dpy-20*. The *mep-1* rescuing cosmid
15 M04B2 is shown in bold.

Figure 1B shows the predicted MEP-1 protein (SEQ ID NO:1). Zinc finger motifs are shaded, and the positions of *mep-1* mutations are indicated by arrowheads.

Figure 2 shows the genomic sequence of *mep-1* (SEQ ID NO:2). The
20 start and stop codons are indicated by highlighting.

Figure 3 shows the nucleic acid sequence of the *mep-1* open reading frame (SEQ ID NO:3).

Figure 4 shows the deduced amino acid sequence of MEP-1.

Figures 5A and 5B are bar graphs showing that *trr-1* single mutants are
25 defective in P(8).p fate specification. Induction of individual P(3-8).p cells was scored in wild-type animals (Figure 5A) and *trr-1(n3712)* mutants (Figure 5B). Certain cells in *trr-1* mutants adopted hybrid fates in which one of two Pn.p daughters divided like daughters of induced Pn.p cells and the other daughter remained undivided as in uninduced Pn.p cells. Ectopic induction in single

mutant animals containing each of the other five *trr-1* mutations was similarly restricted to P8.p.

Figure 6 is a bar graph showing that. *trr-1* and class B synMuv mutations are synthetically defective in P8.p cell-fate specification. P8.p induction was scored. We recognized *trr-1* homozygous mutants as non-Gfp progeny of *trr-1/ mIn1[dpy-10(e128) mIs14]* heterozygous parents. *lin-15B(n744)*, *lin-35(n745)*, *lin-36(n766)* and *lin-37(n758)* are the strongest mutations of their corresponding genes. Strains homozygous for these mutations are viable. *trr-1; synmuvB* double mutant strains with these mutations were derived from parents that were homozygous for the *synmuvB* mutation and hence lacked maternal and zygotic function of the class B synMuv gene in question. The *dpl-1(n3316)* null mutation causes sterility. We combined *dpl-1(RNAi)* with the *dpl-1(n3316)* mutation to generate mutants that lacked both maternal and zygotic *dpl-1* activity and recognized these mutants as non-Gfp progeny of *dpl-1(n3316) trr-1/ mIn1[dpy-10(e128) mIs14]* heterozygous parents that were injected with *dpl-1* dsRNA.

Figure 7A shows the *trr-1* gene structure as derived from cDNA and genomic sequences. Shaded boxes indicate coding sequence and open boxes indicate 5' and 3' untranslated regions. Predicted translation initiation and termination codons and the poly(A) tail are indicated. Positions of alternative splicing are indicated by asterisks. In all cases, the use of alternative splice acceptors creates small differences in the *trr-1* coding sequence: alternative splicings of the fourth (ag/TTTCAGAC (SEQ ID NO:4) versus agtttcag/AC (SEQ ID NO:5)), fifth (ag/AATCTTCAGTC (SEQ ID NO:6) versus (agaatcttcag/CC (SEQ ID NO:7)), eleventh (ag/AACTTTAAGAT (SEQ ID NO:8) versus agaactttaag/AT (SEQ ID NO:9) and twelfth introns (ag/TTGCAGAA (SEQ ID NO:10) versus agttgcag/AA (SEQ ID NO:11)) differ by either six or nine nucleotides.

Figure 7B is a schematic diagram of the TRR-1 protein. The positions of substitutions caused by TRR-1 mutations are indicated above. TRR-1 is

similar to mammalian TRRAP and yeast Tra1p throughout the lengths of the proteins. Domains of similarity (e.g., FAT and ATM/PI-3 kinase-like domains) that these three proteins share are indicated.

Figure 8 shows the genomic nucleic acid sequence of *trr-1* (SEQ ID NO:12). The start and stop codons are indicated by highlighting.

Figure 9 shows the nucleic acid sequence of the *trr-1* open reading frame (SEQ ID NO:13).

Figure 10 shows the deduced amino acid sequence of TRR-1 (SEQ ID NO:14).

Figure 11A is a schematic diagram showing the *hat-1* gene structure as derived from cDNA and genomic sequences. Shaded boxes indicate coding sequence and open boxes indicate 5' and 3' untranslated regions. Predicted translation initiation and termination codons and the poly(A) tail are shown.

Figure 11B is a schematic diagram of the HAT-1 protein. HAT-1 is similar to MYST family acetyltransferases, all of which contain a MOZ/SAS acetyltransferase domain and some of which contain a chromodomain. Nematodes expressing the *hat-1(n4075)* deletion are expected to produce only the first 35 amino acids of the wild-type HAT-1 protein and additional frameshifted amino acids prior to truncation.

Figure 11C is a bar graph showing that *hat-1* single mutants were defective in P(8).p fate specification. Induction of individual P(3-8).p cells was scored in wild-type animals (left) and *hat-1(n4075)* mutants (right). *hat-1* homozygous mutants were recognized as non-Unc progeny of *+/nT1n754*; *hat-1(n4075)/nT1n754* heterozygous parents.

Figure 11D is a bar graph showing that *hat-1* is synthetically defective in P8.p cell-fate specification with the class B synMuv mutation *lin-15B(n744)*. P8.p induction was scored as described below. *hat-1* homozygous mutants were recognized as in (C).

Figure 12 shows the genomic nucleic acid sequence of *hat-1* (SEQ ID NO:15). The start and stop codons are indicated by highlighting.

Figure 13 shows the nucleic acid sequence of the *hat-1* open reading frame (SEQ ID NO:16).

5 Figure 14 shows the deduced amino acid sequence of HAT-1 (SEQ ID NO:17).

Figure 15A is a schematic diagram showing *epc-1* and *ssl-1* gene structures and deletion mutations. The gene structure of *epc-1* was derived by comparing cDNA and genomic sequences.

10 Figure 15B is a schematic showing the *ssl-1* gene structure and deletion mutation. The gene structure of *ssl-1* is partially derived from comparison of cDNA and genomic sequences (SL1 splice leader, 5' untranslated region, exons 1-12 and the beginning of exon 13) and partially predicted solely from genomic sequence (the end of exon 13). As we do not have cDNA clones representing
15 the 3' end of *ssl-1*, we are unable to reliably assign a 3' untranslated region and poly(A) tail. Filled boxes indicate coding sequence and open boxes indicate 5' and 3' untranslated regions. SL1 splice leaders, predicted translation start and stop codons and poly(A) tail are shown. The regions of genomic sequence removed by the *epc-1*(n4076) and *ssl-1*(n4077) deletions are indicated.

20 Figure 16 shows the genomic nucleic acid sequence of *epc-1* (SEQ ID NO:18).

Figure 17 shows the nucleic acid sequence of the *epc-1* open reading frame (SEQ ID NO:19).

25 Figure 18 shows the deduced amino acid sequence of EPC-1 (SEQ ID NO:20).

Figure 19 shows the genomic nucleic acid sequence of *ssl-1* (SEQ ID NO:21) and the deduced amino acid sequence.

Figure 20A shows the exon boundaries of the *ssl-1* genomic nucleic acid sequence.

Figure 20B shows the cDNA nucleic acid sequence of *ssl-1* (SEQ ID NO:22).

Figure 21 shows the amino acid sequence of SSL-1 (SEQ ID NO:23).

Figures 22A and 22B are schematic diagrams showing two models of TRR-1/HAT-1/EPC-1 function with respect to class B synMuv proteins

Figure 22A is a schematic diagram showing that a TRR-1/HAT-1/EPC-1 complex and the class B synMuv proteins act on different targets and differentially regulate transcription. In this model a putative TRR-1/HAT-1/EPC-1 complex acts on targets that are different from those of a putative class B synMuv protein complex. A TRR-1/HAT-1/EPC-1 complex may promote transcription of genes that negatively regulate vulval development, whereas class B synMuv proteins may repress transcription of genes that promote vulval development.

Figure 22B is a schematic diagram showing a second model. In this second model, a TRR-1/HAT-1/EPC-1 complex acts on the same targets as do the class B synMuv proteins. Together these two putative protein complexes may specify an acetylation pattern on histones that is required for efficient silencing of genes that promote vulval development. A TRR-1/HAT-1/EPC-1 complex may act through DPL-1 and EFL-1, although genetic interactions suggest that not all TRR-1/HAT-1/EPC-1 complex activity goes through DPL-1 and EFL-1.

Figure 23 shows the genomic sequence of *lin(n3628)* including 1 kb of upstream and downstream genomic sequences (SEQ ID NO:24). The exon boundaries are also defined.

Figure 24 shows the amino acid sequence of LIN(n3628) (SEQ ID NO:25).

Figure 25 shows the genomic sequence of *lin(n4256)* (SEQ ID NO:26). The exon boundaries are also defined.

Figure 26 shows the amino acid sequence of LIN(n4256) (SEQ ID NO:27).

Figure 27 shows the genomic sequence of *lin-65* (SEQ ID NO:28). The exon boundaries are also defined.

Figure 28 shows the amino acid sequence of LIN-65 (SEQ ID NO:29). The exon boundaries are also defined.

5 Figure 29 shows the mRNA sequence that encodes the LIN(n3628) human ortholog, KIAA1732.

Figure 30 shows the amino acid sequence of KIAA1732 (SEQ ID NO:35).

10 Figure 31 defines the domains of LIN(n3628), including the SET catalytic domain.

Figure 33 defines the domains of KIAA1732, including the SET catalytic domain.

Description of the Invention

15 As reported in more detail below, we have identified new components of the Rb pathway that function in chromatin remodeling and antagonize Ras signaling, and methods for using such components for the identification of chemotherapeutics and the identification of new clinical targets for the treatment of neoplasia.

20

Example I

Isolation of new synMuv mutants

A variety of genetic studies revealed that sterility is often associated with a severe reduction of class B synMuv gene function. For example, in a genetic screen for alleles that did not complement the synMuv phenotype of *lin-9(n112)*, (Ferguson et al., *Genetics* 123: 109-21, 1989) recovered the alleles *lin-9(n942)* and *lin-9(n943)*, which caused sterility when homozygous. In another example, we performed gene dosage studies and observed that, in comparison to the wild-type *lin-52(n771)/Df* and *dpl-1(n2994)/Df* heterozygotes had markedly reduced brood sizes. In addition, deletion mutations of synMuv genes that showed recessive sterility were recovered by reverse genetic approaches (e.g. alleles of *lin-53* (LU 1999), *lin-54*, and *dpl-1* (Ceol et al., *Mol Cell* 7: 461-73, 2001).

Previous genetic screens for synMuv mutants (Ferguson et al., *Genetics* 123: 109-21, 1989) were performed before a link between loss of synMuv gene function and sterility was well established. These screens required that isolates be fertile and viable in order to recover mutant alleles. In addition to failing to recover recessive sterile mutations of the genes described above, these screens failed to recover mutations of the class B synMuv genes *efl-1* and *let-418*, both of which can mutate to a sterile phenotype (Von Zelewsky et al., *Development* 127: 5277-84, 2000; Ceol et al., *Mol Cell* 7: 461-73, 2001). Given this failure, we undertook a genetic screen to identify additional synMuv genes that would allow the recovery of homozygous sterile mutations through phenotypically wild-type heterozygous siblings.

To screen for new synMuv mutants, we examined the F₂ progeny of individually plated F₁ animals after EMS mutagenesis of *lin-15A(n767)* mutants. This screen represented 6760 haploid genomes examined for mutations that either alone or in combination with *lin-15A(n767)* showed a recessive Muv phenotype. Using this strategy we identified 95 Muv mutations, 24 of which were maintained as heterozygotes due to recessive sterility that

co-segregated with the Muv phenotype. Three mutations caused a Muv phenotype in the absence of *lin-15A*(n767) and were found to affect *lin-1* and *lin-31*, both of which function downstream of *let-60* Ras in vulval induction (Ferguson et al., *Nature* 326:259-67, 1987). These mutations, *lin-1*(n3443),
5 *lin-1*(n3522), and *lin-31*(n3440) were not characterized further. Additionally, we recovered 29 mutations that, together with *lin-15A*(n767), caused a weakly penetrant (< 30%) Muv phenotype. The remaining 63 mutations were assigned to 21 complementation groups, which include the previously known genes *ark-1*, *dpl-1*, *efl-1*, *gap-1*, *let-418*, *lin-9*, *lin-13*, *lin-15B*, *lin-35*, *lin-36*, *lin-52*,
10 *lin-53*, *lin-61*, and *sli-1*, and the new genes *lin*(n3441), *lin*(n3542), *lin*(n3628), *lin*(n3681), *lin*(n3707), *mep-1*, and *trr-1*.

Phenotypes of new mutants

We characterized the penetrance of the Muv phenotype for each strain at
15 15°C and 20°C. The results of this study are described in Table 1.

Table 1 Penetrance of Muv phenotype (n)

Genotype	15° C	20° C	Additional phenotypes
<i>ark-1(n3524) lin-15A(n767)</i>	0 (251)	80 (171)	
<i>ark-1(n3701); lin-15A(n767)</i>	12 (190)	95 (160)	
<i>dpl-1(n3643); lin-15A(n767)</i>	99 (154)	100 (252)	
<i>efl-1(n3639); lin-15A(n767)</i>	93 (74)	100 (78)	Ste
<i>gap-1(n3535) lin-15A(n767)</i>	1.4 (143)	50 (236)	
<i>let-418(n3536); lin-15A(n767)</i>	0 (201)	55 (183)	hs Ste
<i>let-418(n3626); lin-15A(n767)</i>	1.6 (62)	97 (76)	Ste
<i>let-418(n3629); lin-15A(n767)</i>	0 (52)	86 (58)	Ste
<i>let-418(n3634); lin-15A(n767)</i>	0 (87)	92 (48)	Ste
<i>let-418(n3635); lin-15A(n767)</i>	0 (76)	71 (70)	Ste
<i>let-418(n3636); lin-15A(n767)</i>	0 (77)	92 (78)	Ste
<i>let-418(n3719); lin-15A(n767)</i>	0 (101)	100 (60)	Ste
<i>lin-9(n3631); lin-15A(n767)</i>	100 (42)	100 (72)	Ste
<i>lin-9(n3675); lin-15A(n767)</i>	43 (166)	100 (105)	
<i>lin-9(n3767); lin-15A(n767)</i>	100 (67)	100 (56)	Ste
<i>lin-13(n3642); lin-15A(n767)</i>	3.3 (60)	100 (63)	Ste
<i>lin-13(n3673); lin-15A(n767)</i>	61 (145)	97 (129)	
<i>lin-13(n3674); lin-15A(n767)</i>	78 (131)	100 (191)	hs Ste
<i>lin-13(n3726); lin-15A(n767)</i>	31 (225)	99 (149)	hs Ste

Genotype	15° C	20° C	Additional phenotypes
<i>lin-15B(n3436) lin-15A(n767)</i>	100 (193)	100 (212)	
<i>lin-15B(n3676) lin-15A(n767)</i>	18 (167)	72 (130)	
<i>lin-15B(n3677) lin-15A(n767)</i>	99 (111)	100 (122)	
<i>lin-15B(n3711) lin-15A(n767)</i>	100 (186)	100 (156)	
<i>lin-15B(n3760) lin-15A(n767)</i>	32 (171)	100 (150)	
<i>lin-15B(n3762) lin-15A(n767)</i>	63 (113)	97 (116)	
<i>lin-15B(n3764) lin-15A(n767)</i>	96 (232)	100 (199)	
<i>lin-15B(n3766) lin-15A(n767)</i>	55 (132)	100 (173)	
<i>lin-15B(n3768) lin-15A(n767)</i>	80 (159)	100 (302)	
<i>lin-15B(n3772) lin-15A(n767)</i>	100 (220)	100 (191)	
<i>lin-35(n3438); lin-15A(n767)</i>	100 (153)	100 (126)	partial Ste at 20°C, Rup
<i>lin-35(n3763); lin-15A(n767)</i>	100 (108)	100 (160)	partial Ste at 20°C, Rup
<i>lin-36(n3671); lin-15A(n767)</i>	65 (191)	100 (151)	
<i>lin-36(n3672); lin-15A(n767)</i>	98 (198)	100 (178)	
<i>lin-36(n3765); lin-15A(n767)</i>	0 (184)	37 (202)	
<i>lin-52(n3718); lin-15A(n767)</i>	100 (41)	100 (82)	Ste
<i>lin-53(n3448); lin-15A(n767)</i>	67 (130)	100 (211)	partial Ste at 20°C

Genotype	15° C	20° C	Additional phenotypes
<i>lin-53(n3521); lin-15A(n767)</i>	100 (34)	100 (125)	partial Ste at 20°C
<i>lin-53(n3622); lin-15A(n767)</i>	85 (61)	100 (66)	Ste
<i>lin-53(n3623); lin-15A(n767)</i>	24 (55)	100 (51)	Ste
<i>lin-61(n3442); lin-15A(n767)</i>	22 (130)	100 (152)	
<i>lin-61(n3446); lin-15A(n767)</i>	36 (124)	99 (191)	
<i>lin-61(n3447); lin-15A(n767)</i>	11 (121)	87 (207)	
<i>lin-61(n3624); lin-15A(n767)</i>	0 (152)	89 (231)	
<i>lin-61(n3736); lin-15A(n767)</i>	0 (193)	100 (201)	
<i>n3441; lin-15A(n767)</i>	80 (165)	99 (195)	
<i>n3541; lin-15A(n767)</i>	79 (242)	98 (137)	
<i>n3543; lin-15A(n767)</i>	85 (177)	100 (121)	
<i>n3628; lin-15A(n767)</i>	2.9 (103)	84 (188)	
<i>n3681; lin-15A(n767)</i>	0 (214)	72 (192)	
<i>n3542 lin-15A(n767)</i>	0 (127)	35 (218)	
<i>n3707 lin-15A(n767)</i>	3.8 (80)	77 (26)	
<i>mep-1(n3680); lin-15A(n767)</i>	4.9 (122)	97 (105)	hs Ste
<i>mep-1(n3702); lin-15A(n767)</i>	30 (61)	100 (141)	Ste
<i>mep-1(n3703); lin-15A(n767)</i>	25 (72)	100 (107)	Ste
<i>sli-1(n3538) lin-15A(n767)</i>	4.3 (138)	90 (173)	
<i>sli-1(n3544) lin-15A(n767)</i>	4.6 (153)	80 (265)	cs embryonic lethality
<i>sli-1(n3683) lin-15A(n767)</i>	5.0 (80)	88 (148)	cs embryonic lethality
<i>trr-1(n3630); lin-15A(n767)</i>	3.1 (131)	85 (212)	Ste, Gro
<i>trr-1(n3637); lin-15A(n767)</i>	1.1 (92)	80 (200)	Ste, Gro

Genotype	15° C	20° C	Additional phenotypes
<i>trr-1(n3704); lin-15A(n767)</i>	3.1 (96)	79 (244)	Ste, Gro
<i>trr-1(n3708); lin-15A(n767)</i>	2.0 (151)	84 (228)	Ste, Gro
<i>trr-1(n3709); lin-15A(n767)</i>	1.0 (97)	77 (154)	Ste, Gro
<i>trr-1(n3712); lin-15A(n767)</i>	5.8 (121)	77 (192)	Ste, Gro

Ste: sterile; Gro: growth rate abnormal; Rup: rupture at the vulva; cs: cold sensitive; hs: heat sensitive.

The penetrance of the Muv phenotype was determined after growing synMuv mutant strains at the indicated temperature for two or more generations. For most strains in which a fully penetrant sterile phenotype was associated with the Muv phenotype, we scored the penetrance of the Muv phenotype by examining sterile progeny of heterozygous mutant parents. For *trr-1* mutant strains, we scored the penetrance of the Muv phenotype by examining non-Gfp progeny of *trr-1 / mIn1[dpy-10(e128)mIs14]; lin-15A(n767)* heterozygous parents. All strains were backcrossed to *lin-15A(n767)* twice prior to phenotypic characterization. In addition to the phenotypes described above, many of the strains exhibited heat sensitive inviability due to frequent rupture, sterility, and/or general sickness.

The penetrance at 25°C is not shown because all strains had a highly penetrant (>90%) Muv phenotype at this temperature. Since a heat-sensitive Muv phenotype is characteristic of most synMuv strains, including those with null mutations in synMuv genes (Ferguson et al., *Genetics* 123: 109-21, 1989), it is likely that many synMuv mutations are not particularly temperature sensitive, but rather that the synMuv genes regulate a temperature sensitive process.

A subset of our synMuv strains also exhibited a sterile phenotype. In these strains, the sterile phenotype cosegregated with the Muv phenotype during backcrosses and two- and three-factor mapping experiments. For those mutations tested, we found that our new mutations did not complement the sterile phenotypes caused by previously isolated, allelic synMuv mutations. These observations suggest that the sterile and Muv phenotypes of these strains were caused by the same mutation.

We observed an unusual aspect to the sterility of one of our strains. We examined the *mep-1(n3680); lin-15A(n767)* strain and found that its sterile phenotype showed maternal-effect rescue. When derived from heterozygous parents, the sterility of the *mep-1(n3680); lin-15A(n767)* animals was 3.2% penetrant (n=62), but was 55% penetrant (n=69) when these animals were derived from homozygous parents. Mutations that affect the Mes (Mes, maternal-effect sterility) genes also show maternal-effect rescue of sterility (Capowski et al., *Genetics* 129: 1061-72, 1991). Some Mes genes encode homologs of *Drosophila* polycomb group proteins and are proposed to function in X chromosome transcriptional silencing in the germline (Holdeman et al., *Development* 125: 2457-67, 1998; Korf et al., *Development* 125: 2469-78, 1998; Fong et al., *Science* 296: 2235-8, 2002). A functional relationship between the synMuv and Mes genes has not been previously reported.

15 New synMuv genes

Using two-factor crosses and sex chromosome transmission tests, we mapped the new mutations to linkage groups (Table 2).

Table 2 Chromosomal linkages of new synMuv mutations**A. Autosomal mutations**

New mutation	Mutation used for selection of homozygous F₂ hermaphrodites	Genotype of selected F₂ hermaphrodites with respect to the linked, unselected mutation
<i>ark-1(n3524)</i>	<i>dpy-20(e1282) IV</i>	2/19 <i>ark-1(n3524)/+</i>
<i>ark-1(n3701)</i>	<i>ark-1(n3701)</i>	1/14 <i>dpy-20(e1282)/+ IV</i>
<i>dpl-1(n3643)</i>	<i>dpl-1(n3643)</i>	0/20 <i>rol-6(e187)/+ II</i>
<i>efl-1(n3639)</i>	<i>rol-4(sc8) V</i>	4/20 <i>efl-1(n3639)/+</i>
<i>let-418(n3536)</i>	<i>let-418(n3536)</i>	4/21 <i>rol-4(sc8)/+ V</i>
<i>let-418(n3626)</i>	<i>rol-4(sc8) V</i>	0/19 <i>let-418(n3626)/+</i>
<i>let-418(n3629)</i>	<i>rol-4(sc8) V</i>	1/20 <i>let-418(n3629)/+</i>
<i>let-418(n3634)</i>	<i>rol-4(sc8) V</i>	2/19 <i>let-418(n3634)/+</i>
<i>let-418(n3635)</i>	<i>rol-4(sc8) V</i>	5/20 <i>let-418(n3635)/+</i>
<i>let-418(n3636)</i>	<i>rol-4(sc8) V</i>	3/20 <i>let-418(n3636)/+</i>
<i>let-418(n3719)</i>	<i>rol-4(sc8) V</i>	2/30 <i>let-418(n3719)/+</i>
<i>lin-9(n3631)</i>	<i>unc-32(e189) III</i>	0/20 <i>lin-9(n3631)/+</i>
<i>lin-9(n3675)</i>	<i>lin-9(n3675)</i>	0/22 <i>unc-32(e189)/+ III</i>
<i>lin-9(n3767)</i>	<i>lin-9(n3767)</i>	0/16 <i>mgP21/+ III</i>
<i>lin-13(n3642)</i>	<i>unc-32(e189) III</i>	1/20 <i>lin-13(n3642)/+</i>
<i>lin-13(n3673)</i>	<i>lin-13(n3673)</i>	0/25 <i>unc-32(e189)/+ III</i>
<i>lin-13(n3674)</i>	<i>lin-13(n3674)</i>	0/25 <i>unc-32(e189)/+ III</i>
<i>lin-13(n3726)</i>	<i>lin-13(n3726)</i>	1/26 <i>unc-32(e189)/+ III</i>
<i>lin-35(n3438)</i>	<i>lin-35(n3438)</i>	0/30 <i>dpy-5(e61)/+ I</i>
<i>lin-35(n3763)</i>	<i>lin-35(n3763)</i>	0/22 <i>dpy-5(e61)/+ I</i>
<i>lin-36(n3671)</i>	<i>lin-36(n3671)</i>	1/23 <i>unc-32(e189)/+ III</i>
<i>lin-36(n3672)</i>	<i>lin-36(n3672)</i>	0/16 <i>unc-32(e189)/+ III</i>
<i>lin-36(n3765)</i>	<i>lin-36(n3765)</i>	0/9 <i>unc-32(e189)/+ III</i>
<i>lin-52(n3718)</i>	<i>lin-52(n3718)</i>	1/16 <i>mgP21/+ III</i>
<i>lin-53(n3448)</i>	<i>lin-53(n3448)</i>	1/22 <i>dpy-5(e61)/+ I</i>
<i>lin-53(n3521)</i>	<i>dpy-5(e61) I</i>	0/20 <i>lin-53(n3521)/+</i>
<i>lin-53(n3622)</i>	<i>dpy-5(e61) I</i>	5/30 <i>lin-53(n3622)/+</i>
<i>lin-53(n3623)</i>	<i>lin-53(n3623)</i>	4/16 <i>hP4/+ I</i>
<i>lin-61(n3442)</i>	<i>lin-61(n3442)</i>	0/20 <i>dpy-5(e61)/+ I</i>
<i>lin-61(n3446)</i>	<i>lin-61(n3446)</i>	1/23 <i>dpy-5/+ I</i>

New mutation	Mutation used for selection of homozygous F ₂ hermaphrodites	Genotype of selected F ₂ hermaphrodites with respect to the linked, unselected mutation
<i>lin-61(n3447)</i>	<i>lin-61(n3447)</i>	0/13 <i>dpy-5(e61)/+ I</i>
<i>lin-61(n3624)</i>	<i>lin-61(n3624)</i>	0/15 <i>dpy-5(e61)/+ I</i>
<i>lin-61(n3736)</i>	<i>dpy-5(e61) I</i>	1/19 <i>lin-61(n3736)/+</i>
<i>lin(n3441)</i>	<i>lin(n3441)</i>	5/20 <i>dpy-5(e61)/+ I</i>
<i>lin(n3541)</i>	<i>lin(n3541)</i>	9/31 <i>dpy-5(e61)/+ I</i>
<i>lin(n3543)</i>	<i>lin(n3543)</i>	9/27 <i>dpy-5(e61)/+ I</i>
<i>lin(n3628)</i>	<i>lin(n3628)</i>	1/29 <i>dpy-5(e61)/+ I</i>
<i>lin(n3681)</i>	<i>lin(n3681)</i>	3/22 <i>rol-4(sc8)/+ V</i>
<i>mep-1(n3680)</i>	<i>mep-1(n3680)</i>	0/30 <i>dpy-20(e1282)/+, IV</i>
<i>mep-1(n3702)</i>	<i>mep-1(n3702)</i>	0/16 <i>sP4/+ IV</i>
<i>mep-1(n3703)</i>	<i>mep-1(n3703)</i>	0/16 <i>sP4/+ IV</i>
<i>trr-1(n3630)</i>	<i>rol-6(e187) II</i>	0/20 <i>trr-1(n3630)/+</i>
<i>trr-1(n3637)</i>	<i>rol-6(e187) II</i>	1/20 <i>trr-1(n3637)/+</i>
<i>trr-1(n3704)</i>	<i>rol-6(e187) II</i>	1/30 <i>trr-1(n3704)/+</i>
<i>trr-1(n3708)</i>	<i>rol-6(e187) II</i>	0/20 <i>trr-1(n3708)/+</i>
<i>trr-1(n3709)</i>	<i>rol-6(e187) II</i>	2/30 <i>trr-1(n3709)/+</i>
<i>trr-1(n3712)</i>	<i>rol-6(e187) II</i>	1/19 <i>trr-1(n3712)/+</i>

B. X-linked mutations

New mutation	Criteria for X linkage
<i>lin(n3542)</i>	transmission test
<i>lin(n3707)</i>	transmission test
<i>gap-1(n3535)</i>	transmission test
<i>lin-15B(n3436)</i>	males with pseudovulva
<i>lin-15B(n3676)</i>	transmission test, males with pseudovulva
<i>lin-15B(n3677)</i>	males with pseudovulva
<i>lin-15B(n3711)</i>	males with pseudovulva
<i>lin-15B(n3760)</i>	transmission test, males with pseudovulva
<i>lin-15B(n3762)</i>	males with pseudovulva
<i>lin-15B(n3764)</i>	transmission test, males with pseudovulva
<i>lin-15B(n3766)</i>	transmission test, males with pseudovulva
<i>lin-15B(n3768)</i>	transmission test, males with pseudovulva
<i>lin-15B(n3772)</i>	transmission test, males with pseudovulva
<i>sli-1(n3538)</i>	transmission test
<i>sli-1(n3544)</i>	transmission test
<i>sli-1(n3683)</i>	transmission test

Autosomal and sex chromosome linkages were determined as described below. *lin(n3541)* was also mapped relative to *bli-3(e767)* and *unc-54(e1092)*, mutations present on the extreme left and right arms, respectively, of linkage group I. Of 16 Muv progeny selected from a *lin(n3541) / bli-3(e767) unc-54(e1092); lin-15A(n767)* parent, none were *bli-3(e767)/+* whereas six were *unc-54(e1092)/+*, indicating *lin(n3541)* lies nearer to *bli-3(e767)*.

We then determined if a given mutation failed to complement mutations of known synMuv genes on the same linkage group. Mutations that were not assigned to known synMuv complementation groups were tested against unassigned mutations within the same linkage group for complementation. These tests defined seven new synMuv loci: *trr-1*, *mep-1*, *lin(n3441)*, *lin(n3628)*, *lin(n3681)*, *lin(n3707)*, and *lin(n3542)*. We used three-factor

crosses to map most of these new synMuv genes within their respective linkage groups (Table 3).

Table 3 Map data for newly-identified synMuv loci

5

A. Three- and four-factor mapping

Gene	Genotype of heterozygote	Phenotype of selected recombinants	Genotype of selected recombinants (with respect to unselected markers)
<i>ark-1</i>	+ + <i>ark-1</i> / <i>unc-5 dpy-20</i> +; <i>lin-15A</i> (n767)	Unc	10/10 <i>ark-1</i> / +
		Dpy	0/1 <i>ark-1</i> / +
	+ <i>ark-1</i> + / <i>dpy-20</i> + <i>unc-30</i> ; <i>lin-15A</i> (n767)	Dpy	15/35 <i>ark-1</i> / +
		Unc	17/33 <i>ark-1</i> / +
	<i>dpy-20</i> + + <i>ark-1</i> / + <i>lin-3 unc-22</i> +; <i>lin-15A</i> (n767)	Dpy	3/9 <i>unc-22</i> / +
		Muv	3/3 <i>unc-22</i> / +
	<i>dpy-20</i> + <i>ark-1</i> + / + <i>unc-22</i> + <i>unc-30</i> ; <i>lin-15A</i> (n767)	Dpy	1/3 <i>unc-22</i> / +
		Muv	1/2 <i>unc-22</i> / +
		Unc-22	2/3 <i>ark-1</i> / +
		Unc-30	5/6 <i>ark-1</i> / +
	<i>dpy-20</i> + <i>ark-1</i> + / + <i>dpy-26</i> + <i>unc-30</i> ; <i>lin-15A</i> (n767)	Dpy-20	4/7 <i>dpy-26</i> / +
		Muv	3/8 <i>dpy-26</i> / +
<i>gap-1</i>	+ + <i>gap-1 lin-15A</i> (n767) / <i>unc-1 dpy-3</i> + <i>lin-15A</i> (n767)	Unc	17/17 <i>gap-1</i> / +
		Dpy	0/8 <i>gap-1</i> / +
	<i>gap-1</i> + + <i>lin-15A</i> (n767) / + <i>unc-2 lon-2 lin-15A</i> (n767)	Unc	0/2 <i>gap-1</i> / +
		Lon	6/6 <i>gap-1</i> / +
<i>lin-52</i>	+ <i>gap-1</i> + <i>lin-15A</i> (n767) / <i>dpy-3</i> + <i>unc-2 lin-15A</i> (n767)	Unc	14/18 <i>gap-1</i> / +

Gene	Genotype of heterozygote	Phenotype of selected recombinants	Genotype of selected recombinants (with respect to unselected markers)
	<i>+ lin-52 + / unc-16 + unc-47; lin-15A(n767)</i>	Unc-47	7/9 <i>lin-52 / +</i>
	<i>lin-52 + unc-69 / + stP127 +; lin-15A(n767)</i>	Muv	3/12 <i>stP127 / +</i>
	<i>sma-3 + lin-52 + / + sqv-3 + unc-69; lin-15A(n767)</i>	Sma	9/9 <i>sqv-3 / +</i>
		Muv	1/27 <i>sqv-3 / +</i>
		Unc	14/16 <i>lin-52 / +</i>
<i>lin(n3441)</i>	<i>+ lin(n3441) + / bli-3 + lin-17; lin-15A(n767)</i>	Lin-17	9/19 <i>lin(n3441) / +</i>
	<i>bli-3 + lin(n3441) / + spe-15 +; lin-15A(n767)</i>	Muv	10/18 <i>spe-15 / +</i>
	<i>+ lin(n3441) lin-17 / spe-15 + +; lin-15A(n767)</i>	Lin-17	11/11 <i>spe-15 / +</i>
<i>lin(n3628)</i>	<i>lin(n3628) + + / + dpy-5 unc-13; lin-15A(n767)</i>	Dpy	0/6 <i>lin(n3628) / +</i>
		Unc	6/6 <i>lin(n3628) / +</i>
	<i>+ lin(n3628) + / unc-11 + dpy-5; lin-15A(n767)</i>	Unc	1/11 <i>lin(n3628) / +</i>
		Dpy	5/11 <i>lin(n3628) / +</i>
	<i>unc-11 + + lin(n3628) / + unc-73 lin-44 +; lin-15A(n767)</i>	Muv	3/9 <i>unc-73 lin-44 / + +</i>
	<i>+ + lin(n3628) dpy-5 / unc-73 lin-44 + +; lin-15A(n767)</i>	Muv	0/21 <i>unc-73 lin-44 / + +</i>
	<i>lin(n3628) + dpy-5 / + unc-38 +; lin-15A(n767)</i>	Muv	3/7 <i>unc-38 / +</i>
	<i>unc-11 lin(n3628) + / + + unc-38; lin-15A(n767)</i>	Muv	0/9 <i>unc-38 / +</i>
<i>lin(n3542)</i>	<i>+ + + lin(n3542) lin-15A(n767) / unc-10 dpy-6 lin-15A(n767)</i>	Unc	8/8 <i>lin(n3542) / +</i>
	<i>+ lin(n3542) + lin-15A(n767) / dpy-6 + unc-9 lin-15A(n767)</i>	Unc	4/40 <i>lin(n3542) / +</i>
<i>mep-1</i>	<i>+ mep-1 + / unc-5 + dpy-20; lin-15A(n767)</i>	Unc	56/57 <i>mep-1 / +</i>
		Dpy	2/61 <i>mep-1 / +</i>
	<i>mep-1 + + / + dpy-20 unc-30; lin-15A(n767)</i>	Dpy	0/51 <i>mep-1 / +</i>
		Unc	58/58 <i>mep-1 / +</i>
	<i>+ + mep-1 + / unc-24 mec-3 + dpy-20; lin-15A(n767)</i>	UncMec	10/12 <i>mep-1 / +</i>

Gene	Genotype of heterozygote	Phenotype of selected recombinants	Genotype of selected recombinants (with respect to unselected markers)
<i>sli-1</i>		Unc	17/17 <i>mep-1</i> / +
		MecDpy	0/8 <i>mep-1</i> / +
		Dpy	2/8 <i>mep-1</i> / +
	+ <i>mep-1 dpy-20</i> + / <i>lin-3</i> + + <i>unc-22</i> ; <i>lin-15A(n767)</i>	Dpy	5/5 <i>lin-3</i> / +
		Vul	3/10 <i>mep-1</i> / +
	+ + <i>mep-1</i> + / <i>mec-3 sem-3</i> + <i>dpy-20</i> ; <i>lin-15A(n767)</i>	Mec	17/17 <i>mep-1</i> / +
		Dpy	6/13 <i>mep-1</i> / +
	<i>sli-1</i> + + <i>lin-15A(n767)</i> / + <i>lon-2 unc-6 lin-15A(n767)</i>	Lon	0/6 <i>sli-1</i> / +
	<i>sli-1</i> + + <i>lin-15A(n767)</i> / + <i>unc-2 lon-2 lin-15A(n767)</i>	Lon	5/5 <i>sli-1</i> / +
	<i>sli-1</i> + + <i>lin-15A(n767)</i> / + <i>dpy-3 unc-2 lin-15A(n767)</i>	Dpy	0/10 <i>sli-1</i> / +
<i>trr-1</i>		Unc	6/6 <i>sli-1</i> / +
	<i>sli-1</i> + + <i>lin-15A(n767)</i> / + <i>unc-1 dpy-3 lin-15A(n767)</i>	Unc	0/14 <i>sli-1</i> / +
		Dpy	10/10 <i>sli-1</i> / +
	+ <i>rol-6</i> + <i>trr-1</i> / <i>dpy-10</i> + <i>unc-4</i> +; <i>lin-15A(n767)</i>	Rol	3/14 <i>unc-4</i> / +
		Dpy	3/3 <i>trr-1</i> / +
		Unc	0/8 <i>trr-1</i> / +
	+ <i>trr-1</i> + / <i>dpy-10</i> + <i>rol-1</i> ; <i>lin-15A(n767)</i>	Rol	9/20 <i>trr-1</i> / +
	+ + <i>trr-1</i> / <i>dpy-10 unc-53</i> +; <i>lin-15A(n767)</i>	Unc	0/17 <i>trr-1</i> / +
	+ <i>trr-1</i> + / <i>unc-53</i> + <i>rol-1</i> ; <i>lin-15A(n767)</i>	Unc	7/10 <i>trr-1</i> / +
		Rol	7/10 <i>trr-1</i> / +
	+ <i>trr-1</i> + <i>rol-1</i> / <i>unc-4</i> + <i>mex-1</i> +; <i>lin-15A(n767)</i>	Rol	12/14 <i>mex-1</i> / +

B. Deficiency mapping

Gene	Genotype of heterozygote	Phenotype of heterozygote
<i>lin-52</i>		

	<i>unc-36 lin-52 / nDf40 dpy-18; lin-15A(n767)</i>	Muv
<i>mep-1</i>	<i>mep-1 / sDf63 unc-31; lin-15A(n767) / +</i>	PvlSte
	<i>mep-1 / sDf62 unc-31; lin-15A(n767) / +</i>	PvlSte
	<i>mep-1 / sDf10; lin-15A(n767) / +</i>	WT
<i>trr-1</i>	<i>rol-6 trr-1 / mnDf57; lin-15A(n767)</i>	WT
	<i>rol-6 trr-1 / unc-4 mnDf90; lin-15A(n767)</i>	WT
	<i>rol-6 trr-1 / mnDf29; lin-15A(n767)</i>	WT
	<i>trr-1 / unc-4 mnDf87; lin-15A(n767)</i>	Muv

WT: wild-type; Pvl: protruding vulva; Ste: sterile.

Three- and four-factor crosses were performed using standard methods (Brenner, *Genetics* 77: 71-94, 1974). Deficiency heterozygotes were constructed as described below. In addition, we have isolated *trr-1*, *mep-1*, *lin(n3628)*, and *lin(n3681)* mutations away from the parental *lin-15A(n767)* mutation. *mep-1*, *lin(n3628)*, and *lin(n3681)* mutations alone do not cause a Muv phenotype, and *trr-1* mutations alone cause only weak ectopic vulval induction. Thus, these mutations synergize with *lin-15A(n767)* and are indeed synMuv mutations.

We identified mutations in *gap-1* and *sli-1*, two genes that were originally identified in screens for mutations that suppressed the Vul phenotype caused by a reduction in *let-60* Ras pathway signaling (Jongeward et al., *Genetics* 139: 1553-66, 1995; Hajnal et al., *Genes Dev* 11: 2715-28, 1997). We also identified mutations in *ark-1*, a gene that was first identified in a screen for mutations that caused ectopic vulval induction in a *sli-1* mutant background (Hopper et al., *Mol Cell* 6: 65-75, 2000). *gap-1*, *sli-1*, and *ark-1* single mutants were previously isolated and found to have no (*sli-1*, *gap-1*) or subtle (*ark-1*) defects in vulval development. Our results indicate that *sli-1*, *gap-1*, and *ark-1* act redundantly with *lin-15A* to negatively regulate *let-60* Ras signaling.

Molecular identification of *mep-1*

We isolated three mutations, *n3680*, *n3702* and *n3703*, in a gene that we mapped to a small interval on linkage group IV in between *sem-3* and *dpy-20* as shown in Figure 1. We attempted to rescue the Muv phenotype of *n3680*; *lin-15A(n767)* mutants using cosmid clones from this interval. Transgenic animals containing the cosmid M04B2 were rescued for the Muv phenotype and also showed improved fertility relative to non-transgenic animals. The genomic sequence of *mep-1* is shown in Figure 2. The *mep-1* open reading frame sequence is shown in Figure 3. This gene was originally identified based on its interaction with the germline specification genes *mog-1*, *mog-4*, *mog-5* and *pie-1* in yeast two-hybrid screens (Belfiore et al. RNA. 8:725-39, 2002). Because somatic tissues adopt germ cell-specific characteristics in *mep-1* mutants, *mep-1* is thought to repress germ cell fates in the soma. We sequenced *mep-1* in our mutant strains to determine if the mutations we isolated affected this gene. These mutations identify functionally important amino acid residues or domains. *n3680* mutants have a missense mutation that, in the predicted MEP-1 protein, changes a polar serine residue to an asparagine. *n3702* mutants have a nonsense mutation and *n3703* mutants a splice acceptor mutation in the *mep-1* gene. Our genetic mapping data, cosmid rescue, and DNA sequence results indicate that *n3680*, *n3702*, and *n3703* are *mep-1* mutations.

The deduced amino acid sequence of MEP-1 is shown in Figure 4. *mep-1* encodes a protein containing six zinc-finger motifs. Zinc fingers are known to mediate interactions of proteins with DNA and with other proteins. The zinc fingers of MEP-1 likely mediate interactions with LET-418 or other synMuv proteins.

Sequences of synMuv mutations

We determined sequences of mutations that affected additional synMuv genes (Table 4).

Table 4 Selected synMuv proteins and allele sequences**A. Features of selected synMuv proteins**

Protein	No. amino acids	Protein similarities and domains
DPL-1	598	Similar to DP family transcription factors; Contains DNA- and E2F-binding domains
EFL-1	342	Similar to E2F family transcription factors; Contains DNA-binding, DP-binding and transactivation domains
LET-418	1829	Similar to Mi-2 family ATP-dependent chromatin remodeling enzymes; Contains chromodomains, PHD finger motifs and a helicase domain*
LIN-9	LIN-9L: 644 LIN-9S: 642	Similar to <i>Drosophila</i> Aly cell cycle regulator and mammalian proteins of unknown function
LIN-13	2248	Protein has 24 Zn-finger motifs
LIN-35	961	Similar to Retinoblastoma (pRb) family transcriptional regulators; Contains "pocket" interaction domain
LIN-36	962	Novel protein with C/H-rich and Q-rich regions
LIN-52	161	Similar to <i>Drosophila</i> and mammalian proteins of unknown function
LIN-53	417	Similar to <i>Drosophila</i> p55, mammalian RbAp48 subunits of chromatin remodeling and histone deacetylase complexes; Contains WD repeats
LIN-61	491	Similar to <i>Drosophila</i> l(3)mbt and other MBT repeat-containing proteins
MEP-1	853	Protein has six Zn finger motifs
SLI-1	582	Similar to Cbl family ubiquitination-promoting proteins; Contains SH2 domain and RING finger motif
TRR-1	4064 [†]	Similar to mammalian TRRAP transcriptional regulator

B. Allele sequences

Mutation	Wild-type sequence	Mutant sequence	Substitution, splice site change or aberration	Domain affected by missense mutation
<i>dpl-1</i> (n3643)	TAT	TAA	Y341ochre	-
<i>efl-1</i> (n3639)	CAA	TAA	Q175ochre	-
<i>let-</i> <i>418</i> (n3536)	CCT	CTT	P675L	helicase/ATPase
<i>let-</i> <i>418</i> (n3626)	GGT	AGT	G1006S	helicase/ATPase
<i>let-</i> <i>418</i> (n3629)	TCC	TTC	S925F	helicase/ATPase
<i>let-</i> <i>418</i> (n3634)	TGG	TAG	W1128amber	-
<i>let-</i> <i>418</i> (n3635)	CAG	TAG	Q1594amber	-
<i>let-</i> <i>418</i> (n3636)	ACT	TCT	T807S	helicase/ATPase
<i>let-</i> <i>418</i> (n3719)	TGG	TAG	W295amber	-
<i>lin-9</i> (n3631)	CAA	TAA	LIN-9L: Q594ochre	-
<i>lin-9</i> (n3675)	GAT	AAT	LIN-9S: Q592ochre	-
<i>lin-9</i> (n3767)	CAG	TAG	LIN-9L: D305N	none predicted
<i>lin-</i> <i>13</i> (n3642)	CAT	TAT	LIN-9S: D303N	none predicted
<i>lin-</i> <i>13</i> (n3673)	CAG	TAG	LIN-9L: Q509amber	-
<i>lin-</i> <i>13</i> (n3674)	CGA	TGA	LIN-9S: Q507amber	-
<i>lin-</i> <i>13</i> (n3726)	GGA	GAA	H832Y	Zn finger
			Q1988amber	-
			R1250opal	-
			G229E	none predicted

Mutation	Wild-type sequence	Mutant sequence	Substitution, splice site change or aberration	Domain affected by missense mutation
<i>lin-</i> 35(n3763) ^o	<u>G</u> CA	<u>G</u> TA	A555V	Pocket
	TTG AAA	TTG AAA	K594frameshift and	
	AAG	AAA G	truncation after 611a.a.	-
<i>lin-</i> 36(n3671)	<u>C</u> AT	<u>C</u> CT	H284P	C/H-rich region
	<u>G</u> AA	<u>A</u> AA	E424K	none predicted
<i>lin-</i> 36(n3672)	<u>C</u> AG	<u>T</u> AG	Q467amber	-
<i>lin-</i> 36(n3765) [†]	<u>G</u> CT	<u>G</u> TT	A242V	C/H-rich region
<i>lin-</i> 52(n3718)	<u>C</u> AG	<u>T</u> AG	Q31amber	-
<i>lin-</i> 53(n3448)	<u>A</u> GT	<u>A</u> TT	S384I	WD repeat
<i>lin-</i> 53(n3521)	<u>G</u> AA	<u>A</u> AA	E174K	WD repeat
		AAG/atatgtgt		
<i>lin-</i> 53(n3622)	AAG/gtatgtgt	(SEQ ID NO:30)	Exon 1 donor	-
<i>lin-</i> 53(n3623)	<u>T</u> GG	<u>T</u> AG	W337amber	-
		aacttcag/AAT		
<i>lin-</i> 61(n3442)	aacttcag/AAT	(SEQ ID NO:31)	Exon 4 acceptor	-
<i>lin-</i> 61(n3446)	<u>C</u> AA	<u>T</u> AA	Q412ochre	-
<i>lin-</i> 61(n3447)	<u>A</u> GT	<u>A</u> AT	S354N	MBT repeat
<i>lin-</i> 61(n3624)	<u>C</u> CG	<u>T</u> CG	P132S	none predicted

Mutation	Wild-type sequence	Mutant sequence	Substitution, splice site change or aberration	Domain affected by missense mutation
<i>lin-</i> 61(n3736)	<u>TTT</u>	<u>TCT</u>	F247S	MBT repeat
<i>mep-</i> 1(n3680)	<u>AGT</u>	<u>AAT</u>	S309N	none predicted
<i>mep-</i> 1(n3702)	<u>CAG</u>	<u>TAG</u> CTT/ataagttt (SEQ ID	Q706amber	-
<i>mep-</i> 1(n3703)	CTT/gtaagttt	NO:32)	Exon 3 donor	-
<i>sli-1</i> (n3538)	<u>TCA</u>	<u>TTA</u> ttttccaa/AAA (SEQ ID	S305L	SH2
<i>sli-1</i> (n3544)	ttttccag/AAA	NO:33) tttttaa/GAT (SEQ ID	Exon 6 acceptor	-
<i>sli-1</i> (n3683)	tttttag/GAT	NO:34)	Exon 4 acceptor	-
<i>trr-1</i> (n3630)	<u>TGG</u>	<u>TAG</u>	W2064amber	-
<i>trr-1</i> (n3637)	<u>CAG</u>	<u>TAG</u>	Q3444amber	-
<i>trr-1</i> (n3704)	<u>CAA</u>	<u>TAA</u>	Q694ochre	-
<i>trr-1</i> (n3708)	<u>CGA</u>	<u>TGA</u>	R1248opal	-
<i>trr-1</i> (n3709)	<u>CGA</u>	<u>TGA</u>	R2550opal	-
<i>trr-1</i> (n3712)	<u>TGG</u>	<u>TAG</u>	W2505amber	-

In the "Wild-type sequence" and "Mutant sequence" columns, exon and intron sequences are denoted by uppercase and lowercase script, respectively. Nucleotides altered by mutation are underlined.

5 * The predicted LET-418 protein contains a sequence described as a helicase domain. This domain was originally identified in helicases, but has since been found in non-helicase proteins. Many of these proteins share a common ATPase activity, and this domain contains residues that are important for ATP binding and hydrolysis.

† The adenosine inserted by the *lin-35*(n3763) frameshift mutation is not underlined because it is unclear which nucleotide in the adenosine repeat was inserted.

10 ‡ In addition to the missense mutation described, we found an additional mutation associated with *lin-36*(n3765). This mutation, AG/gtaagaagaaaagc to AG/gtaagaagaaaagt, is present in the third intron of *lin-36* and creates a possible splice donor sequence. If this splice donor were used, an inframe ochre (TAA) stop codon would be encountered, truncating the LIN-36 protein after 261 amino acids.

15 § Due to alternative splicing, *trr-1* encodes proteins that range in length between 4051 and 4061 amino acids

DPL-1 and EFL-1 are described by (Ceol et al., *Mol Cell* 7: 461-73, 2001 and (Page et al., *Mol Cell* 7: 451-60, 2001). LIN-9 is described by Beitel et al., *Gene* 254: 253-63, 2000); LIN-13 is

described by Melendez et al., *Genetics* 155: 1127-37, 2000); LIN-35 and LIN-53 are described by (Lu et al., *Cell* 95:981-91, 1998); LIN-36 is described by (Thomas et al., *Development* 126: 3449-59, 1999); and SLI-1 is described by (Yoon et al., *Science* 269: 1102-5, 1995).

5 Most mutations are GC-to-AT transitions that are characteristic of EMS mutagenesis (Anderson, *Methods Cell Biol* pp. 31-58, 1995). Many of these mutations are predicted to truncate the corresponding synMuv proteins. The truncations predicted by *efl-1(n3639)*, *let-418(n3719)*, and *lin-52(n3718)* are particularly severe, and the synMuv and sterile phenotypes caused by these

10 mutations may represent the null phenotypes of these genes. In addition, we found missense mutations that disrupt predicted functional domains of synMuv proteins. For example, *n3536*, *n3626*, *n3629* and one of the two mutations of *n3636* affect the ATPase/helicase domain of LET-418. LET-418 is a member of the Mi-2 family of ATP-dependent chromatin remodeling enzymes (Solari et al.,

15 *Curr Biol* 10: 223-6, 2000; Von Zelewsky et al., *Development* 127: 5277-84, 2000), and the LET-418 missense mutations suggest that LET-418 function is similarly dependent on ATP hydrolysis. At least one mutation affecting the LIN-13 protein, *n3642*, is predicted to disrupt a canonical zinc-finger motif. This missense mutation indicates that at least some of the twenty-four LIN-13

20 zinc fingers are important for its synMuv activity. Missense mutations affecting other synMuv proteins are not as easily linked to the disruption of predicted functional domains. These mutations may provide a useful starting point in identifying functional motifs within synMuv proteins that are not predicted by sequence comparisons.

25

Frequency of mutant isolation

The rate at which we isolated mutations was much higher than that observed in previous synMuv screens: including those 63 mutations described in this study, we recovered one synMuv mutation per 107 haploid genomes

30 screened versus 1/750 (Ferguson et al., *Genetics* 123: 109-21, 1989), 1/400 and 1/667 in previous screens. We believe the reasons for this difference are threefold. First, our screen design allowed the isolation of synMuv mutations

that also caused sterility. Sterile synMuv mutants were observed previously, but because the heterozygous siblings of these mutants were present in a sea of genotypically unrelated animals, the underlying mutations could not be recovered. Second, our parental strain carried the strong class A mutation, *lin-15A(n767)*. The penetrance of a strain's Muv phenotype is dependent on the aggregate strengths of the component synMuv mutations. Therefore, even weak mutations may be identified in a strong synMuv background such as *lin-15A(n767)*. Although we have not formally tested this possibility, we believe that some of the mutations we recovered only weakly affect synMuv activity. Such mutations may not have been recovered in previous screens that were performed in partial loss-of-function synMuv backgrounds. Third, in screening a plate of many F₂ progeny derived from a single F₁ animal, we observed many genotypically identical animals per haploid genome screened. This type of screening likely accounts for our isolation of a number of partially penetrant synMuv mutations. Such mutations may not have been identified in earlier synMuv screens that typically observed fewer genotypically identical animals per haploid genome screened.

Our high rate of recovery indicates many genes can mutate to a synMuv phenotype. Including the ten genes we identified in this study, a total of 25 genes can act redundantly with class A synMuv genes. Many of these genes are represented by one or a few mutant alleles, indicating that screens for synMuv genes are not saturated.

The synMuv genes we identified likely act in different pathways

Class B synMuv mutations synergize with class A synMuv mutations, but not with other class B synMuv mutations. Such genetic behavior led to the hypothesis that class B synMuv genes are part of a single genetic pathway (Ferguson et al., *Genetics* 123:109-21, 1989). In support of this hypothesis, mutations affecting different class B synMuv genes are similarly suppressed by loss-of-function mutations in the *let-23* receptor tyrosine kinase and other

let-60 Ras pathway loss-of-function mutations (Ferguson et al., *Nature* 326:259-67, 1987), a subset of class B synMuv gene products have been shown to interact *in vitro*, and their homologs are known function together in other systems (Lu et al., *Cell* 95: 981-91, 1998; Ceol et al., *Mol Cell* 7: 461-73, 2001). Because we conducted our screen in a class A synMuv background, we anticipated recovering mutations that affected genes of the class B synMuv pathway. In addition to Class B synMuv mutations, our results suggest that we recovered mutations that disable distinct genetic pathways. We recovered six mutations that affect the *trr-1* gene. Unlike typical class B synMuv mutations, *trr-1(n3712)* synergize not only with class A synMuv mutations, but also with class B synMuv mutations. *trr-1(n3712)* single mutants also atypically show ectopic vulval induction. Because of its unusual genetic interactions, we propose that *trr-1* functions in a pathway distinct from the class B synMuv pathway. We also recovered mutations affecting the *sli-1*, *gap-1*, and *ark-1* genes. These genes were previously characterized as negative regulators of *let-60* Ras pathway activity, acting genetically downstream of the *let-23* receptor tyrosine kinase (Jongeward et al., *Genetics* 139: 1553-66, 1995; Hajnal, et al., *Genes Dev* 11: 2715-28 1997; Hopper et al., *Mol Cell* 6: 65-75, 2000). The molecular identities of *sli-1*, *gap-1*, and *ark-1* support their action downstream of *let-23*. *sli-1* encodes a homolog of the c-cbl proto-oncoprotein, which is thought to downregulate receptor tyrosine kinase levels through ubiquitin-mediated degradation (Yoon et al., *Science* 269: 1102-5, 1995; Levkowitz et al., *Mol Cell* 4: 1029-40, 1999). *gap-1* is a member of the GTPase-activating protein family (Hajnal, et al., *Genes Dev* 11: 2715-28 1997). GAPs enhance the catalytic function of Ras family GTPases, thereby facilitating the switch from active GTP-bound to inactive GDP-bound Ras. *ark-1* encodes a predicted cytoplasmic tyrosine kinase that interacts with the SEM-5 SH2/SH3 adaptor protein (Hopper et al., *Mol Cell* 6: 65-75, 2000). Since *sem-5* acts downstream of the *let-23* receptor tyrosine kinase, *ark-1* is proposed to inhibit *let-60* Ras signaling downstream of *let-23*. These genetic

and molecular data suggest that *sli-1*, *gap-1*, and *ark-1* directly regulate *let-60* Ras pathway members and are likely not part of the canonical class B synMuv pathway, which is thought to regulate the *let-60* Ras pathway either upstream of, or in parallel to, the *let-23* receptor tyrosine kinase. We are currently
5 placing our synMuv mutations into different genetic classes by examining interactions with class B synMuv and *let-23* mutations.

***lin-52* encodes a new putative Rb pathway protein**

lin-35, a member of the class B synMuv pathway, encodes a protein
10 similar to the mammalian tumor suppressor pRb (Lu et al., *Cell* 95: 981-91, 1998). Other genes with class B synMuv activity encode DP, E2F, RbAp48, histone deacetylase and HP1 family proteins (Lu et al., *Cell* 95: 981-91, 1998; Ceol et al., *Mol Cell*, 7: 461-73, 2001; Couteau et al., *EMBO Rep* 3: 235-41, 2002). Mammalian homologs of these proteins are known to functionally, and
15 in some cases physically, interact with pRb. These and other parallels indicate that the class B synMuv pathway is an analog of Rb pathways in other systems. Consequently, additional class B synMuv genes may have homologs with analogous functions in other systems. One such gene is *lin-52*. By the genetic criteria outlined above, *lin-52* is a class B synMuv gene. *lin-52* mutations
20 synthetically interact with class A mutations, but not with class B mutations. Furthermore, preliminary experiments indicate that the Vul phenotype of a *let-23* loss-of-function mutation is epistatic to the Muv phenotype caused by *lin-52* and *lin-15A* loss of function. *lin-52* encodes a small protein, portions of which are conserved in similarly small proteins predicted by the human, mouse
25 and *Drosophila* genome sequences. The characterization of these and other class B synMuv protein homologs should help to determine whether they too function in Rb-mediated signaling.

The experiments described above were carried out as follows

Strains and general techniques

Strains were cultured as described by (Brenner, *Genetics* 77: 71-94, 1974). and grown at 20°C unless otherwise indicated. The wild-type parent of all the strains described in this study was the *Caenorhabditis elegans* Bristol strain N2. For some two and three-factor mapping experiments we used the polymorphic strain RW7000 (Williams et al., *Genetics* 131: 609-24, 1992). We also used strains containing the following mutations:

LGI: *bli-3(e767)*, *lin-17(n677)*, *unc-11(e47)*, *unc-73(e936)*, *lin-44(n1792)*,
 10 *unc-38(x20)*, *dpy-5(e61)*, *lin-35(n745)*, *lin-61(sy223)*, *unc-13(e1091)*,
lin-53(n833) (Ferguson et al., *Genetics* 123: 109-21 (1989), *unc-54(e1092)* (Dibb et al., *J. Mol Biol* 183: 543-51, 1985).

LGII: *lin-31(n301)*, *dpy-10(e128)*, *tra-2(q276)*, *rol-6(e187)*, *dpl-1(n2994)*,
unc-4(e120), *unc-53(n569)*, *mex-1(it9)*, *rol-1(e91)*

15 LGIII: *dpy-17(e164)*, *lon-1(e185)*, *sma-3(e491)*, *lin-13(n770)* (Ferguson et al., *Genetics* 123: 109-21 (1989), *lin-37(n758)*, *lin-36(n766)*, *unc-36(e251)*,
lin-9(n112), *unc-32(e189)*, *unc-16(e109)*, *sqv-3(n2842)*, *lin-52(n771)* (Ferguson et al., *Genetics* 123: 109-21 (1989), *unc-47(e307)*, *unc-69(e587)*,
dpy-18(e364)

20 LGIV: *lin-1(e1275)*, *unc-5(e53)*, *unc-24(e138)*, *mec-3(e1338)*, *lin-3(n378)*,
sem-3(n1900), *dpy-20(e1282)*, *unc-22(e66)*, *dpy-26(n198)*, *unc-31(e169)*,
unc-30(e191), *lin-54(n2231)*, *dpy-4(e1166)* LGV: *tam-1(cc567)* (Hsieh et al., *Genes Dev* 13: 2958-70, 1999), *unc-46(e177)*, *let-418(s1617)*, *dpy-11(e224)*,
rol-4(sc8), *unc-76(e911)*, *efl-1(n3318)* Ceol et al., *Mol Cell* 7: 461-73 (2001).

25 *dpy-21(e428)* LGX: *sli-1(sy143)*, *aex-3(ad418)*, *unc-1(e1598n1201)*,
dpy-3(e27), *gap-1(gal33)* (Hajnal et al., *Genes Dev* 11: 2715-28, 1997),
unc-2(e55), *lon-2(e678)*, *unc-10(e102)*, *dpy-6(e14)*, *unc-9(e101)*, *unc-3(e151)*,
lin-15A(n767), *lin-15AB(n765)*. Unless otherwise noted, the mutations used are described by (Riddle et al., *C. elegans II* (Cold Spring Harbor, New York, Cold Spring Harbor Laboratory Press 1997). In addition, we used strains

30

containing the following chromosomal aberrations: *mnDf57 II* (Sigurdson, et al., *Genetics* 108: 331-45, 1984), *mnDf90 II* (Sigurdson, et al., *Genetics* 108: 331-45, 1984), *mnDf29 II* (Sigurdson, et al., *Genetics* 108: 331-45, 1984), *mnDf87 II* (Sigurdson, et al., *Genetics* 108: 331-45, 1984),

5 *mIn1[dpy-10(e128)mIs14] II* (Edgley et al., *Mol Genet Genomics* 266: 385-95, 2001), *mnC1[dpy-10(e128) unc-52(e444)] II* (Herman, *Genetics* 88: 49-65, 1978), *nDf40 III* (Hengartner et al., *Nature* 356: 494-9, 1992), *qC1[dpy-19(e1259)glp-1(q339)] III* (Austin, et al., *Cell* 58: 565-571, 1989), *sDf63 IV*, *sDf62 IV* (Clark et al., *Mol Gen Genet* 232: 97-105, 1992), *sDf10 IV*

10 (Rogalski et al., *Genetics* 102: 725-36, 1982), *eT1(III;V)* (Rosenbluth et al., *Genetics* 99: 415-28, 1981), *nT1(IV;V)* (Ferguson et al., *Genetics* 110: 17-72, 1985). *mIs14*, an integrated transgene linked to the chromosomal inversion *mIn1*, consists of a combination of GFP-expressing transgenes that allow

15 *mIs14*-containing animals to be scored beginning at the 4-cell stage of embryogenesis (Edgley et al., *Mol Genet Genomics* 266: 385-95, 2001).

Isolation of new alleles

We mutagenized *lin-15A(n767)* hermaphrodites with ethyl methanesulfonate (EMS) as described by (Brenner, *Genetics* 77: 71-94, 1974).

20 We allowed these animals to recover on food for between 15 minutes to one hour, and then transferred individual P₀ larvae in L4 lethargus to 50 mm plates. After three to five days, 20 F₁ L4 larvae per P₀ were individually transferred to 50 mm plates, and, subsequently, F₂ animals on these plates were screened for a Muv phenotype. We screened the progeny of 3380 F₁ animals using this

25 procedure.

Linkage group assignment

We used the following markers to determine linkage of newly isolated synMuv mutations to autosomes: *dpy-5 I*, *rol-6 II*, *unc-32 III*, *dpy-20 IV*, *rol-4*

30 *V*. We generated animals heterozygous for the new synMuv mutation and for

at least two of these markers. For fertile synMuv mutants we picked Muv progeny and determined if these progeny segregated the markers, whereas for sterile synMuv mutants we picked single marker homozygotes and determined if these animals segregated the synMuv mutation. We also mapped some mutations using polymorphisms present in the RW7000 strain. We generated animals heterozygous for the new synMuv mutation and for RW7000 markers. We picked individual Muv progeny of these animals, performed lysis and used the resulting template DNA to monitor linkage to each of the autosomes by PCR (Williams et al., *Genetics* 131: 609-24, 1992). We tested for sex linkage to assign some new synMuv mutations to the X chromosome. Briefly, we generated heterozygous or hemizygous mutant males and mated them with marked *lin-15A(n767)* hermaphrodites. We then determined whether all, indicating sex linkage, or roughly half, indicating autosomal linkage, of the cross progeny hermaphrodites of this mating segregated the synMuv mutation. Some *lin-15B* mutations were not tested for sex linkage. Instead, we tentatively assigned X-chromosome linkage based on the presence, when *lin-15A(n767)* males were mated with these mutants, of cross-progeny males with pseudovulval ventral protrusions. Such protrusions are often observed in hemizygous *lin-15AB* mutant males (Ferguson et al., *Genetics* 110: 17-72, 1985) but are found at a much lower penetrance in *lin-15A(n767)* males that are hemizygous for an X-linked synMuv mutation affecting genes other than *lin-15B*. The mutations we assigned in this manner were later determined by complementation tests to affect *lin-15B*.

25 Complementation tests

We typically performed complementation tests by mating males heterozygous for the new mutation and hemizygous for *lin-15A(n767)*, or, if X-linked, males hemizygous for both the new mutation and *lin-15A(n767)*, into marked synMuv mutant hermaphrodites, all of which contained a *lin-15A* mutation. Hemizygous *lin-15B(n3711)lin-15A(n767)* males could not mate.

To perform complementation tests with this mutation, we mated *tra-2(q276); lin-15B(n3711)lin-15A(n767)/++* XX males into marked *lin-15AB* hermaphrodites. For new mutations that caused recessive sterility, we generated heterozygous males by starting matings with wild-type L4 males and individual gravid, putative heterozygous mutant hermaphrodites. For complementation tests we used cross-progeny males derived from plates that had self-progeny Muv animals present. In all complementation tests, unmarked cross-progeny hermaphrodites were scored.

10 Construction of deficiency heterozygotes.

To construct *trr-1(n3712)* heterozygotes with *mnDf57*, *mnDf90* and *mnDf29*, *Df/mIn1*; *lin-15A(n767)* males were generated. These males were mated into *rol-6 trr-1(n3712)/mIn1*; *lin-15A(n767)* hermaphrodites and non-Rol, non-Gfp cross-progeny were scored. *mnDf87* heterozygous males do not mate so in this case we generated *lin(n3712)/mnDf87*; *lin-15A(n767)* animals by mating *lin(n3712)/mIn1*; *lin-15A(n767)* males into *unc-4 mnDf87/mIn1*; *lin-15A(n767)* hermaphrodites. To construct the *lin-52* heterozygote with *nDf40*, we mated *nDf40 dpy-18/unc-36*; *lin-15A(n767)* males into *unc-36 lin-52(n771)*; *lin-15A(n767)* hermaphrodites and scored non-Unc cross-progeny. *mep-1/Df* animals were constructed by mating *Df/nT1*; *+/nT1* males into *dpy-20 mep-1*; *lin-15A(n767)* hermaphrodites and scoring non-Dpy cross-progeny.

Transgenic animals

25 Germline transformation was performed, as described by (Mello et al., *Embo J* 10: 3959-70, 1991), by injecting cosmid (5-10 ng/μL) or plasmid (50-80 ng/μL) DNA into *lin-52* or *mep-1* mutants. Either pRF4, which causes a dominant Rol phenotype, or pPD93.97, which expresses *gfp* under the control of the *myo-3* promoter, was used as a coinjection marker.

30

***lin-52* cDNA isolation**

We obtained a partial *lin-52* cDNA clone, yk253b12, that included 249 nucleotides of the *lin-52* open reading frame and also included the 3' untranslated region and a polyA tail. We used the 5' RACE system v2.0 for rapid amplification of chromosome ends (GIBCO-BRL, LIFE TECHNOLOGIES, Inc. Gaithersburg, Maryland) to determine the 5' end of the *lin-52* transcript. We ligated the two portions of the *lin-52* cDNA together to generate a full-length cDNA clone. The *lin-52* 5' RACE products were trans-spliced to the SL2 leader sequence consistent with observations made by (Zorio et al., *Nature* 372: 270-2, 1994).

Allele sequence

We used PCR-amplified regions of genomic DNA as templates in determining gene sequences. For each gene investigated, we determined the sequences of all exons and splice junctions. Whenever observed, the sequence of a mutation was confirmed using an independently-derived PCR product. All sequences were determined using an automated ABI 373 DNA sequencer.

Example II

As detailed below, we have identified a distinct class of genes, termed the class C synMuv genes, that negatively regulate vulval induction.

Proper vulval development in the nematode *C. elegans* requires that specific ectodermal cells, termed Pn.p cells, adopt different cell fates. The specification of Pn.p cells that eventually make vulval tissue occurs in two steps, each of which involves the selection of a subset of Pn.p cells from a larger Pn.p field (Sulston, *Dev Biol* 56: 110-56, 1977). In the first step, which occurs in the L1 larval stage shortly after the Pn.p cells are generated, anterior and posterior Pn.p cells fuse with the syncytial hypodermis. After this first step, the unfused midbody P(3-8).p cells each have the capacity to adopt a vulval cell fate (Sternberg et al., *Cell* 44: 761-72, 1986). In a second step,

however, only three of these cells, P(5-7).p, adopt such fates in which they undergo three rounds of division to generate seven or eight descendants. P3.p, P4.p and P8.p adopt non-vulval fates, typically dividing only once to generate two descendants that eventually fuse with the syncytial hypodermis. The decision to adopt vulval cell fates occurs during the L2 and early L3 larval stages and is followed by cell divisions and differentiation in the L3 and L4 larval stages, respectively (Sternberg et al., *Cell* 44: 761-72, 1986; Ferguson et al., *Nature* 326: 259-67, 1987). While mutations in class C synMuv genes alone cause mild defects, when a class C gene mutation is combined with either a class A or class B mutation, the two mutations synergize to produce more severe vulval induction and other developmental defects. Class C synMuv genes, *trr-1*, *hat-1*, and *epc-1*, encode homologs of the transcriptional coactivator TRRAP, the MYST family acetyltransferases TIP60 and Esa1p and the *Drosophila* Enhancer of Polycomb (E(Pc)) protein, respectively. Because of the predicted acetyltransferase activity of the HAT-1 protein and because orthologs TRRAP and E(Pc) family proteins have been copurified in histone acetyltransferase complexes, we propose that a combination of histone acetyltransferase and histone deacetylase activities is required to properly specify vulval cell fates in *C. elegans*.

20

***trr-1* interacts with class A and class B synMuv mutations**

We performed a genetic screen for synMuv mutants in a *lin-15A*(n767) background and identified six mutations in our pool of isolates that failed to complement each other and that defined the gene *trr-1*. To quantitate the synMuv phenotype in these mutants, we scored the number of cells that were induced to become vulva.

To more precisely quantitate the Muv phenotype of *trr-1*; *lin-15A* strains, we scored the numbers of P(3-8).p cells induced per animal and found that all strains had a similarly penetrant, temperature-sensitive hyperinduced phenotype (Table 5A).

30

Table 5 *trr-1* mutations cause a hyperinduced phenotype

A. <i>trr-1</i> interactions with synMuv mutations				
Genotype	Temp (°C)	Ave. # P(3-8).p induced (\pm SE)	% animals hyperinduced	n
wild-type	20	3.00 (\pm 0)	0	31
<i>lin-15A(n767)</i>	20	3.00 (\pm 0)	0	24
<i>lin-38(n751)</i>	20	3.00 (\pm 0)	0	27
<i>trr-1(n3630); lin-15A(n767)</i>	20	4.52 (\pm 0.15)	82	45
<i>trr-1(n3637); lin-15A(n767)</i>	20	4.52 (\pm 0.14)	83	54
<i>trr-1(n3704); lin-15A(n767)</i>	20	4.20 (\pm 0.13)	79	43
<i>trr-1(n3708); lin-15A(n767)</i>	20	4.71 (\pm 0.14)	92	36
<i>trr-1(n3709); lin-15A(n767)</i>	20	4.81 (\pm 0.13)	95	39
<i>trr-1(n3712); lin-15A(n767)</i>	20	4.07 (\pm 0.12)	74	54
<i>lin-15A(n767); trr-1(RNAi)</i>	20	5.60 (\pm 0.08)	100	44
<i>trr-1(n3712) lin-38(n751)</i>	20	4.14 (\pm 0.23)	79	14
<i>lin-38(n751); trr-1(RNAi)</i>	20	5.66 (\pm 0.08)	100	32
wild-type	15	3.00 (\pm 0)	0	29
<i>lin-15A(n767)</i>	15	3.00 (\pm 0)	0	32
<i>trr-1(n3704); lin-15A(n767)</i>	15	3.13 (\pm 0.05)	21	24
<i>trr-1(n3712); lin-15A(n767)</i>	15	3.06 (\pm 0.03)	13	32
wild-type	25	3.00 (\pm 0)	0	36
<i>lin-15A(n767)</i>	25	3.02 (\pm 0.02)	3.6	28
<i>trr-1(n3704); lin-15A(n767)</i>	25	5.87 (\pm 0.06)	100	38
<i>trr-1(n3712); lin-15A(n767)</i>	25	5.47 (\pm 0.14)	100	17

B. *trr-1* single mutants

Genotype	Temp (°C)	Ave. # P(3-8).p induced (\pm SE)	% animals	
			hyperinduced	n
wild-type	20	3.00 (\pm 0)	0	31
<i>trr-1</i> (n3630)	20	3.03 (\pm 0.02)	6.1	33
<i>trr-1</i> (n3637)	20	3.08 (\pm 0.04)	13	30
<i>trr-1</i> (n3704)	20	3.01 (\pm 0.01)	2.6	39
<i>trr-1</i> (n3708)	20	3.05 (\pm 0.03)	8.1	37
<i>trr-1</i> (n3709)	20	3.03 (\pm 0.02)	6.3	32
<i>trr-1</i> (n3712)	20	3.10 (\pm 0.03)	13	89
<i>trr-1</i> (RNAi)	20	3.09 (\pm 0.05)	13	32
wild-type	15	3.00 (\pm 0)	0	29
<i>trr-1</i> (n3704)	15	3.08 (\pm 0.05)	12	26
<i>trr-1</i> (n3712)	15	3.06 (\pm 0.03)	12	25
wild-type	25	3.00 (\pm 0)	0	36
<i>trr-1</i> (n3704)	25	3.04 (\pm 0.03)	3.9	51
<i>trr-1</i> (n3712)	25	3.07 (\pm 0.03)	13	48

The number of P(3-8).p cells induced was scored as described below.

Induction was scored after raising strains at the indicated temperature for two generations. *trr-1* mutant homozygotes were scored by examining the non-Gfp progeny of *trr-1/mIn1[dpy-10(e128) mIs14]* heterozygous parents.

The hyperinduction we observed occurred in P3.p, P4.p and P8.p to similar extents. To determine if *trr-1* interacted with other class A synMuv genes, we constructed a *trr-1*(n3712) *lin-38* double mutant. These double mutant animals were also hyperinduced (Table 5A), suggesting that *trr-1* functions in parallel not only to *lin-15A*, but to the class A synMuv pathway in general.

We also isolated *trr-1*(n3712) and the other *trr-1* mutations away from any other synMuv mutations. Nearly all class A and class B synMuv single mutants adopt a wild-type pattern of P(3-8).p fates (Table 5B), however *trr-1* adults had a weakly penetrant hyperinduced phenotype (Table 5B). By

examining the cell fates adopted by individual P(3-8).p cells in L4 animals, we determined that the vulval cell-fate transformations of *trr-1* single mutants always occurred in P8.p (Figure 5). In addition to ectopic vulval cell-fate transformations, all *trr-1* mutations caused slow growth and sterility, although
5 some mutant animals occasionally produced a small number of eggs (<10, as compared to ~300 for the wild-type), all of which died during embryogenesis.

To determine if *trr-1* interacts with class B synMuv genes, we constructed double mutant strains containing *trr-1(n3712)* and mutations of class B synMuv genes. Interestingly, double mutant strains combining
10 *trr-1(n3712)* with mutations of *lin-15B*, *lin-35* Rb, and *lin-37* showed a significant increase in the penetrance of P8.p transformation (Figure 6). In addition to the increase in P8.p transformation, we occasionally observed ectopic transformations of P3.p and P4.p. Since *lin-15B(n744)*, *lin-35(n745)* and *lin-37(n758)* are strong loss-of-function and possibly null mutations of
15 their corresponding genes, these results indicate that *trr-1* functions redundantly with at least a subset of class B synMuv genes.

No significant increase was observed in *trr-1(n3712); lin-36(n766)* double mutants (Figure 6). By various genetic criteria, this loss-of-function *lin-36* mutation behaves unlike mutations in other class B synMuv genes
20 (Hsieh et al., *Genes Dev* 13: 2958-70, 1999; Fay et al., *Genes Dev* 16: 503-17, 2002). There are at least two possibilities to explain the unusual behavior of *lin-36(n766)*. First, the lack of enhancement could be allele specific, with the *lin-36(n766)* mutation disrupting a function that is redundant with a class A synMuv function but not disrupting a separable *lin-36* function that is
25 redundant with *trr-1* activity. Alternatively, our observations with *lin-36* could reflect a gene-specific lack of enhancement. For example, the strength of the *lin-36* defect may not be equivalent to that of other class B synMuv gene defects such that lack of *lin-36* activity may be readily observable in a class A synMuv background but, unlike other class B synMuv defects, not observable

in a *trr-1* background. Enhancement tests using additional *lin-36* alleles will help to resolve this issue.

***trr-1* encodes a protein similar to mammalian TRRAP**

5 We mapped *trr-1* to a small region of LGII and cloned the gene using transformation rescue as detailed below. To confirm the identity of *trr-1*, we obtained a partial cDNA and, using RNA derived from this cDNA, found that RNA-mediated interference (RNAi) of this gene caused a highly penetrant hyperinduced phenotype in *lin-15A* and *lin-38* mutant backgrounds (Table 5).
10 As determined by RT-PCR and 5' RACE, the *trr-1* gene consists of 22 exons, four of which are alternatively spliced (Figure 7A). Since the sites of alternative splicing are separated by only six or nine nucleotides, the most exclusive (4054 amino acids) and inclusive (4064 amino acids) isoforms differ slightly in size. The genomic sequence of *trr-1* is shown in Figure 8. The
15 sequence of the *trr-1* open reading frame is shown in Figure 9.

 The deduced amino acid sequence of TRR-1 is shown in Figure 10. The predicted TRR-1 proteins are similar to mammalian myc-associated protein TRRAP (transformation/transcription domain-associated protein) and its yeast homolog Tra1p throughout most of their lengths (McMahon et al., *Cell* 94:
20 363-74, 1998; McMahon et al., *Cell* 94: 363-74, 1998; Saleh et al., *J Biol Chem* 273: 26559-65, 1998). TRRAP and Tra1p are similarly large proteins, extending 3828 and 3744 amino acids, respectively. The largest predicted TRR-1 isoform is 25 percent identical to TRRAP and 19 percent identical to Tra1p. TRR-1, TRRAP, and Tra1p share limited regions of homology with
25 other proteins (Figure 7B). One of these regions is located at the carboxy terminus and is similar to the catalytic domains of ATM and PI-3-like kinases. Interestingly, the DXXXXN (SEQ ID NO:29) and DFG motifs critical for kinase activity are not present in TRR-1, TRRAP, or Tra1p (Hunter et al., *Cell* 83: 1-4, 1995). Instead of having an enzymatic function, this domain of
30 TRRAP has been proposed to mediate protein-protein interactions (McMahon

et al., *Cell* 94: 363-74, 1998). All six *trr-1* mutations introduce nonsense codons (Figure 7B). *trr-1(n3637)* is predicted to truncate the protein just prior to the ATM/PI-3 kinase-like domain. The phenotypic strength of *trr-1(n3637)* is similar to that of other alleles, suggesting that deletion of the ATM/PI-3
 5 kinase-like domain alone results in a severe loss of protein function. Finally, *trr-1(n3630)*, *trr-1(n3637)*, and *trr-1(n3712)* introduce amber stop codons, and we observed that the sterility associated with these alleles was reduced by the *sup-5(e1464)* informational suppressor tRNA mutation. This suppression, along with the partially penetrant sterility caused by *trr-1(RNAi)*, confirms that
 10 the sterility observed in *trr-1* mutants is truly due to loss of *trr-1* function.

***trr-1(RNAi)* is synthetically lethal with mutations in *lin-35* Rb and other class B synMuv genes**

trr-1(RNAi) caused more severe phenotypic consequences than did *trr-1*
 15 mutations. For example, the ectopic induction phenotype of *lin-15A*; *trr-1(RNAi)* mutants was much stronger than that of *trr-1*; *lin-15A* mutant strains (Table 5). We do not believe this difference is reflective of a partial loss of gene function caused by all of the *trr-1* mutations. Instead we propose that at least some of the mutations cause a severe loss of gene function and that the
 20 difference is due to an effect of *trr-1(RNAi)* on maternally-provided gene activity. In support of this proposal, *trr-1(n3704)/mnDf87*; *lin-15A* and *trr-1(n3712)/mnDf87*; *lin-15A* mutants that were severely deficient in zygotically-provided *trr-1* activity but retained maternally-provided *trr-1* activity had phenotypic penetrances that were similar to those of *trr-1*; *lin-15A*
 25 homozygotes and were weaker than those of *lin-15A*; *trr-1(RNAi)* mutants. Also arguing that *trr-1*; *lin-15A* homozygotes have significantly reduced zygotically-provided *trr-1* gene activity, the protein truncations predicted by *trr-1(n3704)* and other *trr-1* mutations are likely to remove functional domains and compromise TRR-1 activity.

We further characterized the effects of *trr-1(RNAi)*. In wild-type and class A synMuv genetic backgrounds, *trr-1(RNAi)* caused retarded growth, adult sterility and weakly penetrant embryonic and larval lethalties (Table 6).

Table 6 *trr-1(RNAi)* is synthetically lethal with class B but not with class A synMuv mutations

Genotype	% dead embryos	% dead L1 larvae	Total % lethality
			(n)
wild-type	0	0	0 (1062)
<i>trr-1(RNAi)</i>	6.6	1.2	7.8 (726)
<i>lin-15A(n767)</i>	0	0	0 (823)
<i>lin-38(n751)</i>	0.1	0	0.1 (1003)
<i>lin-15B(n744)</i>	0.2	0	0.2 (1002)
<i>lin-35(n745)</i>	0.6	0.2	0.8 (482)
<i>lin-36(n766)</i>	0.3	0	0.3 (890)
<i>dpl-1(n2994)</i>	14	1.1	15.1 (265)
<i>lin-15A(n767); trr-1(RNAi)</i>	3.2	0.9	4.1 (470)
<i>lin-38(n751); trr-1(RNAi)</i>	3.8	1.3	5.1 (628)
<i>lin-15B(n744); trr-1(RNAi)</i>	62.5	36.0	98.5 (469)
<i>lin-35(n745); trr-1(RNAi)</i>	66.2	33.8	100 (263)
<i>lin-36(n766); trr-1(RNAi)</i>	19.4	21.6	41.0 (444)
<i>dpl-1(n2994); trr-1(RNAi)</i>	45.1	53.6	98.7 (304)

Animals injected with *trr-1* dsRNA were individually plated 10-15

- 5 hours following injection. Injected animals were subsequently transferred to new plates every 24 hours until egg laying had ceased. Dead embryos and larvae on a plate were counted at least two days after eggs were laid. All of the mutant strains in which *trr-1(RNAi)* was performed are homozygous viable.

- 10 Interestingly, *trr-1(RNAi)* caused highly penetrant embryonic and larval lethalties in combination with many class B synMuv mutations. Most of the dead embryos arrested at the late embryonic pretzel stage and those that

hatched died shortly thereafter. We have not yet determined a basis for this lethality. It is important to note that many of the class B synMuv mutations tested are predicted to have severe effects on their cognate class B synMuv proteins. Since *trr-1(RNAi)* can synthetically interact with strong reduction-of-
5 function or null class B synMuv mutations, these data indicate that *trr-1* functions redundantly with class B synMuv genes not only in vulval cell-fate determination but also in an essential process earlier in development.

trr-1(RNAi) causes synthetic lethality in a *lin-36(n766)* background although the penetrance of this lethality is not as high as in other class B
10 synMuv backgrounds. This assay therefore unmasks a redundancy between *trr-1* and *lin-36* that we did not observe in the P8.p induction assay. As discussed above, the strength of the *lin-36* defect may not be equivalent to the strengths of defects of other class B synMuv genes. This difference in strengths may explain why, relative to other class B synMuv genes, *lin-36*
15 shows weaker interactions with *trr-1* in terms of synthetic lethality and synthetic P8.p induction.

***trr-1* synthetically interacts with *dpl-1* DP**

Mammalian TRRAP and yeast Tra1p are thought to function as
20 coactivator proteins that bridge transcription factors to histone acetyltransferases (McMahon et al., *Cell* 94: 363-74, 1998; Brown et al., *Science* 292, 2333-7, 2001). Based on coimmunoprecipitation and functional assays, E2F transcription factors were linked to TRRAP (McMahon et al., *Cell* 94: 363-74, 1998; Lang et al., *J Biol Chem* 276: 32627-34, 2001). *In vivo* E2F
25 and DP family proteins form heterodimers that are bound by Rb family proteins via a direct interaction with the E2F subunit reviewed by (Dyson, *Genes Dev* 12: 2245-62, 1998; Trimarchi et al., *Nat Rev Mol Cell Biol* 3: 11-20, 2002). We previously determined that one of two *C. elegans* E2F family members, *efl-1*, and the sole DP family member, *dpl-1*, are class B synMuv genes Ceol et al., *Mol Cell* 7: 461-73 (2001). As noted above, *lin-35* Rb was also
30

characterized as a class B synMuv gene, and the LIN-35 Rb protein was found to form a complex with DPL-1 and EFL-1 *in vitro* (Lu et al., *Cell* 95: 981-91, 1998; Ceol et al., *Mol Cell* 7: 461-73, 2001).

LIN-35 Rb and Rb proteins in other species are thought to recruit histone
 5 deacetylase complexes to regulate E2F-dependent transcription
 (Brehm et al., *Nature* 391: 597-601, 1998; (Luo et al., *Cell* 92, 463-73, 1998; Magnaghi-Jaulin et al., *Nature* 391: 601-5, 1998). Coupling these results with our genetic finding that *trr-1* acts redundantly with *lin-35* Rb to negatively
 10 regulate vulval induction, one might speculate that EFL-1 and DPL-1 recruit distinct LIN-35-containing and TRR-1-containing complexes to appropriately regulate vulval cell fate determination. To examine this possibility, we wished to determine if *trr-1* acted through *efl-1* and *dpl-1* to negatively regulate vulval development.

Without being tied to a particular theory, three lines of evidence suggest
 15 that *trr-1* does not act solely through transcription factors, *efl-1* and *dpl-1*; first, the ectopic induction of P8.p in *dpl-1 trr-1* double mutants is greater than that observed in either single mutant (Figure 6). Because of the sterility conferred by the *dpl-1*(*n3316*) null and *trr-1*(*n3712*) mutations, these mutants were derived from *dpl-1*(*n3316*) *trr-1*(*n3712*) / ++ mothers. It is notable that in this
 20 test we substantially reduced maternally-provided *dpl-1* activity by injecting mothers with *dpl-1* dsRNA and scoring *dpl-1*(*n3316 RNAi*) *trr-1*(*n3712*) progeny; second, in a weak *lin-15A* mutant background at 15°C, *trr-1*(*RNAi*) greatly enhanced the ectopic induction observed in *dpl-1* mutant animals that were derived from *dpl-1* heterozygous mutant mothers (Table 7);

25

Table 7 *trr-1* acts redundantly with *dpl-1*

Genotype	Ave. # P(3-8).p induced	
	(±SE)	% animals mutant (n)
<i>lin-15A</i> (<i>n433</i>); <i>trr-1</i> (<i>RNAi</i>)	3.17 (±)	20 (15)
<i>dpl-1</i> (<i>n3316</i>); <i>lin-15A</i> (<i>n433</i>)	3.00 (±0)	0 (35)

*dpl-1(n3316); lin-15A(n433);*4.98 (\pm)

89 (45)

trr-1(RNAi)

Animals were raised at 15°C, a temperature at which *dpl-1(n3316); lin-15A(n433)* mutants do not show hyperinduction. *dpl-1(n3316)* homozygous mutants were recognized as the Unc non-Gfp progeny of *dpl-1(n3316) unc-4(e120)/mIn1[dpy-10(e128) mIs14]* heterozygous parents.

5 third, when performed in a homozygous *dpl-1* mutant background, *trr-1(RNAi)* caused synthetic lethality with *dpl-1* (Table 6). Since viable *trr-1(RNAi) dpl-1* progeny could be derived from heterozygous, but not homozygous *dpl-1* mutant mothers, this synthetic lethality apparently required a lack of maternally-provided *dpl-1* activity. These results indicate that *trr-1* does not
 10 act only through *dpl-1* to regulate vulval development and embryonic and larval viability. Although all of these assays were conducted in *dpl-1* mutant backgrounds, we expect that, since reduction of *dpl-1* function is predicted to affect all *C. elegans* DP/E2F activity, these results similarly apply to *efl-1*.

In addition to these data, one other observation argues against the model
 15 that *trr-1* acts solely through *dpl-1*. Whereas double mutants containing *lin-35(n745)*, a putative null allele of *lin-35*, and *trr-1(n3712)* display highly penetrant ectopic induction of P8.p, the ectopic induction in *dpl-1(n3316 RNAi)* mutants is relatively weak (Figure 6). If both *lin-35* and *trr-1* were acting solely through *dpl-1*, defects of equivalent strengths would be expected.

20

The Muv phenotype of *trr-1* mutants requires *let-60* Ras pathway activity

Previous studies determined that a conserved Ras pathway induces vulval development in *C. elegans* reviewed by (Sternberg et al., *Trends Genet* 14: 466-72, 1998). Loss-of-function mutations affecting genes in this pathway
 25 cause a vulvaless (Vul) phenotype characterized by P(3-8).p adopting hypodermal instead of vulval cell fates. To determine if Ras pathway activity is required for the *trr-1* mutant phenotype, we constructed strains in which the functions of *trr-1*, *lin-15A* and a Ras pathway gene were reduced. The uninduced phenotype caused by *let-23* receptor tyrosine kinase and *let-60* Ras

mutations was epistatic to the hyperinduced phenotype caused by *ttr-1* and *lin-15A* loss of function (Table 8).

Table 8 *trr-1* epistasis with *let-23* RTK, *let-60* Ras and *lin-3* EGF

Genotype	Ave. # P(3-8).p induced (\pm SE)	% animals hyperinduced	n
wild-type	3.00 (\pm 0)	0	31
<i>lin-15A(n767)</i>	3.00 (\pm 0)	0	24
<i>lin-15A(n767); trr-1(RNAi)</i>	5.60 (\pm 0.08)	100	44
<i>let-23(sy97); lin-15A(n767)</i>	0.02 (\pm 0.02)	0	28
<i>let-23(sy97); lin-15A(n767); trr-1(RNAi)</i>	0.05 (\pm 0.03)	0	42
<i>let-60(n1876); lin-15A(n767)</i>	0 (\pm 0)	0	17
<i>let-60(n1876); lin-15A(n767); trr-1(RNAi)</i>	0 (\pm 0)	0	23
<i>lin-3(n378); lin-15A(n767)</i>	0.30 (\pm 0.07)	0	40
<i>lin-3(n378); lin-15A(n767); trr-1(RNAi)</i>	4.35 (\pm 0.20)	85	20

5 *let-23(sy97)* homozygous mutants were recognized as Rol Unc non-Gfp progeny of *rol-6(e187) let-23(sy97) unc-4(e120)/mIn1[dpy-10(e128) mIs14]; lin-15A(n767)* heterozygous parents, and *let-60(n1876)* homozygous mutants were recognized as Unc progeny of *let-23(n1876) unc-22(e66)/nT1; +/nT1; lin-15A(n767)* heterozygous parents.

These results indicate that Ras pathway activity is required to produce the *trr-1; lin-15A* Muv phenotype. By contrast, *trr-1; lin-3; lin-15A* triple mutants showed a wild-type level of induction in P(5-7).p and ectopic induction in P3.p, P4.p and P8.p. *lin-3* encodes an EGF-like protein that is produced by the gonadal anchor cell and is thought to act non-cell autonomously to stimulate Ras pathway activity in P(5-7).p (Hill et al., *Nature* 358: 470-6, 1992).. These findings suggest that a basal level of *lin-3*-independent Ras pathway activity, when combined with mutations in *trr-1* and *lin-15A*, is sufficient to induce vulval cell fates in P(3-8).p.

hat-1 and *epc-1*, but not *ssl-1*, loss of function phenocopies *trr-1*

TRRAP and Tra1p are components of protein complexes that acetylate histones (Allard et al., *Embo J* 18: 5108-19, 1999; reviewed by Brown et al., *Trends Biochem Sci* 25:15-9, 2000). These complexes are distinguished by

their histone acetyltransferase subunits: the mammalian TFTC and p/CAF and the yeast SAGA complexes contain Gcn5 family acetyltransferases, whereas the mammalian TIP60 and the yeast NuA4 complexes contain MYST family acetyltransferases.

5 To determine if TRR-1 might function with a histone acetyltransferase in *C. elegans*, we used RNA-mediated interference to inactivate such genes. Whereas inactivation of a *Gcn5* homolog *Y47G6A.6* had no effect, inactivation of a MYST family gene we have named *hat-1* produced a highly penetrant Muv phenotype in a *lin-15A* background. To further characterize *hat-1*, we
10 isolated a deletion allele, *n4075*, that removes 1010 base pairs from the *hat-1* locus and is predicted to produce a protein that contains the first 35 amino acids of HAT-1 followed by 52 unrelated amino acids prior to termination (Figure 11A). The genomic nucleic acid sequence of *hat-1* is shown in Figure 12. The nucleic acid sequence of the *hat-1* open reading frame is shown in Figure 13.
15 The predicted full-length HAT-1 protein is 458 amino acids long, and this deletion is expected to remove the conserved chromodomain and acetyltransferase catalytic domain (Figure 11B). The amino acid sequence of the wild-type HAT-1 protein is shown in Figure 14. *hat-1(n4075)* mutants exhibited the same spectrum of phenotypes and genetic interactions as *trr-1*
20 mutants. *hat-1(n4075)* single mutants were slow growing and sterile. In combination with class A synMuv mutations, *hat-1(n4075)* caused a severe Muv phenotype characterized by P3.p, P4.p and P8.p ectopic induction (Table 8). Alone, *hat-1(n4075)* caused ectopic induction of P8.p (Figure 11C). In combination with a *lin-15B* mutation, the penetrance of this ectopic induction
25 was greatly increased (Figure 11D).

 The TIP60 and NuA4 complexes contain other proteins in addition to MYST family acetyltransferases. We inactivated *C. elegans* genes encoding homologs of these proteins and identified *epc-1* as a negative regulator of vulval induction. The genomic sequence of *epc-1* is shown in Figure 16. The
30 nucleic acid sequence of the *epc-1* open reading frame is shown in Figure 17.

epc-1 encodes a homolog of the *Drosophila* Enhancer of Polycomb (E(Pc)) protein and similarly named mammalian and yeast proteins. The deduced amino acid sequence of EPC-1 is shown in Figure 18. Aside from their association with MYST family histone acetyltransferases, little is known about the molecular interactions of E(Pc)-like proteins. Inactivation of *epc-1* caused fully penetrant embryonic lethality in the broods of animals injected with RNA. To study the effects of *epc-1* inactivation during postembryonic development, we injected *epc-1* RNA into RNAi-deficient hermaphrodites and subsequently mated these animals with RNAi-competent males, a procedure referred to as “zygotic RNAi” (Herman, *Development* 128: 581-90, 2001). For many genes that act during multiple stages of development, this scheme has been shown to provide sufficient gene activity for embryonic functions, but inadequate gene activity for postembryonic functions. *epc-1(RNAi)* performed in this manner did not affect vulval induction in wild-type animals, but produced a Muv phenotype in *lin-15A* and *lin-38* mutant backgrounds (Table 9).

Table 9 *hat-1* and *epc-1* but not *ssl-1* loss of function phenocopies *trr-1* loss of function

Genotype	Ave. # P(3-8).p	% animals	
	induced (\pm SE)	mutant	n
wild-type	3.00 (\pm 0)	0	31
<i>lin-15A(n767)</i>	3.00 (\pm 0)	0	24
<i>lin-38(n751)</i>	3.00 (\pm 0)	0	27
<i>lin-15B(n744)</i>	3.00 (\pm 0)	0	20
<i>hat-1(n4075)</i>	3.15 (\pm 0.08)	15	20
<i>hat-1(n4075); lin-15A(n767)</i>	3.76 (\pm 0.14)	76	25
<i>hat-1(n4075); lin-15B(n744)</i>	3.71 (\pm 0.10)	77	31
<i>rde-1/+; epc-1(RNAi)</i>	3.00 (\pm 0)	0	65
<i>rde-1/+; lin-15A(n767); epc-1(RNAi)</i>	3.32 (\pm 0.10)	36	33
<i>lin-38(n751); rde-1/+; epc-1(RNAi)</i>	3.29 (\pm 0.02)	31	65
<i>rde-1/+; lin-15B(n744); epc-1(RNAi)</i>	3.03 (\pm 0.02)	4.2	48

<i>rde-1/+; ssl-1(RNAi)</i>	3.00 (± 0)	0	37
<i>rde-1/+; lin-15A(n767); ssl-1(RNAi)</i>	3.00 (± 0)	0	42
<i>rde-1/+; lin-15B(n744); ssl-1(RNAi)</i>	3.01 (± 0.01)	2.9	70

hat-1(n4075) homozygous mutants were recognized as the non-Unc progeny of *+/nT1n754; hat-1(n4075)/nT1n754* heterozygous parents. Since RNAi of *epc-1* and *ssl-1* using standard methods causes highly penetrant embryonic lethality, we performed "zygotic RNAi" as described below.

- 5 A low percentage of P8.p induction was observed in a *lin-15B* background. We recently obtained a deletion allele that removes 886 bases from the *epc-1* locus, including the third and fourth *epc-1* exons (Figure 5A). If the second exon were spliced to the fifth exon, a 137 amino acid protein would be produced that contains the first 109 amino acids of the 795 amino acid
- 10 predicted EPC-1 protein. Preliminary studies indicate that *epc-1(n4076)* homozygotes are sterile and, with respect to vulval induction, show genetic interactions similar to those of *epc-1(RNAi)*, *trr-1* and *hat-1* mutants.

TRRAP copurified with the p400 protein as part of the mammalian TIP60 and p400 complexes (Fuchs et al., *Cell* 106: 297-307, 2001). The p400

15 complex was isolated based on its interaction with the adenovirus E1A oncoprotein and was also shown to associate with c-myc. The p400 protein itself is a member of the SWI2/SNF2 family of proteins, and, like many SWI2/SNF2 family members, was shown to possess ATPase activity. We identified a *C. elegans* homolog of p400, which we named *ssl-1* (*ssl*,

20 SWI2/SNF2-like). *ssl-1* genomic sequence and the predicted SSL-1 protein product are shown in Figure 19. Figure 16B shows the nucleotide positions of the predicted exons with respect to *ssl-1* genomic sequence. The cDNA sequence of *ssl-1* is shown in Figure 20. The deduced protein sequence is shown in Figure 21. The function of *ssl-1* was studied by RNAi. *ssl-1(RNAi)*

25 caused an embryonic lethal phenotype reminiscent of that caused by *epc-1(RNAi)*. In both cases, dead embryos generally arrested just prior to morphogenesis and apparently lacked the hypodermal ridge that is a characteristic of enclosed embryos. We are currently characterizing this phenotype further. "Zygotic" RNAi of *ssl-1*, using the same procedure as

described above, caused no vulval defects in wild-type, *lin-15A*, or *lin-15B* genetic backgrounds. These results suggest that *ssl-1* may act with *epc-1* in an essential embryonic process.

5 ***trr-1* acts redundantly with *lin-35* Rb to antagonize *let-60* Ras signaling**

Identifying factors involved in cell fate determination is important for understanding how cells that contain the same genomic information can adopt different cell fates during animal development. As they help to distinguish P3.p, P4.p and P8.p from P(5-7).p, *trr-1*, *hat-1*, and *epc-1* are such cell fate
10 determination genes. Given their molecular identities, *trr-1*, *hat-1*, and *epc-1* likely act at the level of transcription, either in an instructive or permissive fashion, to create differences in gene expression in P3.p, P4.p and P8.p as compared to P(5-7).p.

Many of the pathways involved in regulating cell fate determination are
15 conserved. In many cases, pathways that control cell fate determination in model organisms has been shown to regulate cellular proliferation in mammals. Pathways that regulate vulval cell fate specification in *C. elegans* provide clear examples. A conserved *let-60* Ras pathway induces vulval cell fates, and this pathway is antagonized by the class B *lin-35* Rb pathway. *trr-1*, and likely *hat-1*
20 and *epc-1*, act in parallel to *lin-35* Rb to negatively regulate *let-60* Ras pathway signaling. These comparisons suggest that mammalian counterparts of *trr-1*, *hat-1*, and *epc-1* may similarly act in parallel to Rb and antagonize Ras in the control of cell proliferation.

25 ***trr-1*, *hat-1*, and *epc-1* likely share a common function**

The vulval phenotypes and genetic interactions of *trr-1*, *hat-1*, and *epc-1* mutants are strikingly similar. In light of the copurification of their mammalian and yeast counterparts, these data strongly suggest that TRR-1, HAT-1, and EPC-1 proteins function as part of a protein complex. To
30 conclusively demonstrate such an interaction, strains containing mutations in

two of these genes will be constructed. If these mutants are acting in the same complex, one would not expect to observe synergism in double mutants. In addition, protein-protein interaction studies will be performed. This complex containing putative complex members, *trr-1*, *hat-1*, and *epc-1* were the only
5 candidates we identified by RNAi. It is possible that these three genes encode an indispensable core of a putative HAT complex that associates with other proteins whose functions are dispensable for proper vulval development. The large size of TRR-1 may require it to be divided into fragments to perform protein interaction studies.

10

***hat-1* mutants likely have defects in histone acetylation**

The best studied MYST family acetyltransferases are the yeast Esa1p and mammalian TIP60 proteins. Esa1p was found to preferentially acetylate histone H4 (Smith et al., *Proc Natl Acad Sci USA* 95: 3561-5, 1998; Clark et
15 al., *Mol Cell Biol* 19: 2515-26, 1999; Suka et al., *Mol Cell* 8: 476-9, 2001) Furthermore, depletion of Esa1p resulted in global reduction of the acetylation of H4 and, to a lesser extent, of other nucleosomal histones (Reid et al., *Mol Cell* 6, 1297-307, 2000; Suka et al., *Mol Cell* 8: 476-9, 2001). HAT-1 function is assayed using commercially available antisera that specifically recognize
20 acetylated isoforms of histones to determine whether *hat-1* mutants have gross defects in histone acetylation. Differences in acetylation between *hat-1* mutants and wild-type animals is determined by whole-mount staining of fixed animals or by chromatin immunoprecipitation.

25 Putative HAT complex function

Histone acetyltransferases have been characterized as transcriptional coactivators (reviewed by Roth et al., *Biochem* 70:81-120, 2001), and TRRAP and its yeast homolog Tra1p are proposed to bridge interactions between activation domains of DNA-binding transcription factors and histone
30 acetyltransferases (Brown et al., *Science* 292, 2333-7, 2001). Therefore, a

putative TRR-1/EPC-1/HAT-1 complex may function in transcriptional activation (Figure 22). If so, one would expect it to activate genes that negatively regulate vulval development.

While most data support the link between acetylation and activation, additional observations suggest that at least some histone acetylation may be important for gene silencing. For example, loss-of-function mutations that affect the MYST family acetyltransferases Sas2p and Sas3p cause defects in silencing of mating type loci and telomeres in yeast (Reifsnyder et al., *Nat Genet* 14:42-9, 1996; Ehrenhofer-Murray et al., *Genetics* 145:923-34, 1997). Sas2p and Sas3p are proposed to acetylate newly-deposited nucleosomes, and the modified acetyllysine residues they create are thought to be important for establishing silencing following DNA replication (Meijsing et al., *Genes Dev* 15: 3169-82, 2001; Osada et al. *Genes Dev* 15:3155-68, 2001). These residues may include acetyllysine 16 on histone H4, which is implicated in mating type loci and telomeric silencing in yeast (Johnson et al., *Embo J* 11: 2201-9, 1992; Meijsing et al., *Genes Dev* 15: 3169-82, 2001). Other acetylated histone isoforms are prevalent in silent chromatin. For instance, *Drosophila* heterochromatin is enriched in acetyllysine 12 of histone H4 (Turner et al., *Cell* 69: 375-84, 1992). Just as a MYST family histone acetyltransferase is linked to silencing, loss-of-function studies in *Drosophila* indicate a role for E(Pc) in transcriptional repression. *E(Pc)* mutations synergize with polycomb group mutations to strongly derepress homeobox genes and act alone as suppressors of variegation to derepress genes that are juxtaposed to heterochromatin (Sato et al., *Genetics* 105: 357-70, 1983; Sinclair et al., *Genetics* 148: 211-20, 1998). These observations allow us to consider the possibility that HAT-1, in association with TRR-1 and EPC-1, may normally downregulate transcription (Figure 22). By this model, one would expect a putative TRR-1/EPC-1/HAT-1 complex to silence genes that are required for vulval cell fates. Because we do not know the relevant targets of TRR-1/EPC-1/HAT-1, we cannot distinguish between transcriptional activating versus repressing models at this time.

Putative TRR-1/EPC-1/HAT-1 complex DNA targeting

Their coimmunoprecipitation and cooperation in reporter gene activation suggest that mammalian TRRAP can be targeted by E2F proteins to DNA (McMahon et al., *Cell* 94: 363-74, 1998; (Lang et al., *J Biol Chem* 276: 32627-34, 2001). We investigated the possibility of TRR-1 targeting by DP/E2F heterodimers by studying genetic interactions between *trr-1* and *dpl-1*. *dpl-1* is the only DP family member in *C. elegans* and therefore loss of *dpl-1* activity is expected to effectively reduce all DP/E2F heterodimer function in the organism. *dpl-1* synthetically interacted with *trr-1* in vulval induction and viability assays. It is especially relevant that we observed synergism in some of these assays when using *dpl-1(n3316 RNAi)* mutants, which are severely compromised for *dpl-1* function. These results combined with the observation that the defects of *trr-1* single mutants are stronger than those of *dpl-1* single mutants suggest that *trr-1* acts only partially or not at all through *dpl-1*. If not only through DPL-1, how might a putative TRR-1/EPC-1/HAT-1 complex be targeted to DNA? Studies in yeast indicate that the TRRAP homolog Tra1p directly interacts with acidic activation domains of transcription factors (Brown et al., *Trends Biochem Sci* 25: 15-9, 2000). TRR-1 may similarly be targeted to DNA by transcription factors other than DPL-1. The assays we have used to characterize *trr-1* provide a means of identifying and evaluating candidate transcription factors and other proteins that may function with TRRAP family members in targeted histone acetylation.

The experiments described in Example II were carried out as described below.

Strains and genetics

Strains were cultured as described by (Brenner, *Genetics* 77: 71-94, 1974), and maintained at 20°C unless otherwise specified. Bristol N2 was used as the wild-type strain. The following mutations were used: LGI: *lin-35(n745)*; LGII: *dpy-10(e128)*, *let-23(sy97)*, *rol-6(e187)*, *dpl-1(n2994, n3316)* (Chapters

- 2, 3), *unc-4(e120)*, *trr-1(n3630, n3637, n3704, n3708, n3709, n3712)* (This study), *mex-1(it9)*, *lin-38(n751)*; LGIII: *lon-1(e185)*, *sup-5(e1464)*, *lin-36(n766)*, *lin-37(n758)*; LGIV: *lin-3(n378)*, *let-60(n1876)* (Beitel et al., *Nature* 348: 503-9, 1990); LGV: *dpy-11(e224)*, *rde-1(ne219)*
- 5 (Tabara et al., *Cell* 99: 123-32, 1999); LGX: *lin-15B(n744)*, *lin-15A(n767, n433)* (Ferguson et al., *Genetics* 123: 109-21, 1989) and, unless otherwise noted, are described in (Riddle et al., *C. elegans II* (Cold Spring Harbor, New York, Cold Spring Harbor Laboratory Press, 1997). The deficiencies *mnDf90* and *mnDf87* (Sigurdson, et al., *Genetics* 108: 331-45, 1984), translocation *nT1*
- 10 *n754* (IV;V) (Ferguson et al., *Genetics* 110: 17-72, 1985), and chromosomal inversion *mIn1[dpy-10(e128) mIs14]* (Edgley et al., *Mol Genet Genomics* 266:385-95, 2001), were also used. *mIs14*, an integrated transgene linked to the chromosomal inversion *mIn1*, consists of a combination of GFP-expressing transgenes that allow *mIs14*-containing animals to be identified
- 15 beginning at the 4-cell stage of embryogenesis (Edgley et al., *Mol Genet Genomics* 266:385-95, 2001).

P(3-8).p induction assay

- In the wild-type, P(5-7).p adopt vulval fates in which they divide during
- 20 the L3 larval stage to generate seven or eight descendants. P3.p, P4.p and P8.p adopt non-vulval fates, typically dividing once to generate two descendants that fuse with the hypodermis. Induction was scored in L4 hermaphrodites using Nomarski DIC microscopy by counting the number of descendants produced by individual P(3-8).p cells. Different scores, 1, 0.5 and 0 cells induced, were
- 25 assigned to cells that were fully, partially or not induced, respectively. Partially induced P(3-8).p cells have one daughter that produces a complement of induced descendants while the other daughter fails to divide.

***trr-1* cloning**

We mapped *trr-1* to an interval on LGII between the right endpoint of the deficiency *mnDf90* and the *mex-1* gene. To clone the *trr-1* gene, we performed transformation rescue as described by (Mello et al., *Embo J* 10: 3959-70, 1991), using the pRF4 plasmid (80 ng/μL) as a coinjection marker. We rescued the *trr-1* Muv and sterile phenotypes by injecting the cosmid C47D12 (10ng/μL) into *trr-1(n3712)/mIn1[dpy-10(e128) mIs14]; lin-15A(n767)* mutants and isolating Rol non-Gfp transgenic lines. *trr-1* corresponds to the predicted gene *C47D12.1*.

RNAi analyses

Templates for *in vitro* transcription reactions were made by PCR amplification of either cDNAs and their flanking T3 and T7 promoter sequences or coding exons from genomic DNA using T3 and T7-tagged oligonucleotides. *In vitro*-transcribed RNA was annealed and injected as described by (Fire et al., *Nature* 391: 806-11, 1998).

In addition to the genes described above, we injected RNA corresponding to *C. elegans* genes that encode homologs of the TRRAP complex proteins TIP48/TAP54α (*C. elegans* predicted gene *T22D1.1*), TIP49/TAP54 (*C27H6.2*), Eaf3p (*Y37D8A.9*), p33ING (*Y51H1A.4*), and AF-9 (*M04B2.3*) (Loewith et al., *Mol Cell Biol* 20: 3807-16, 2000; Eisen et al., *J Biol Chem* 276: 3484-91, 2001; Fuchs et al., *Cell* 106: 297-307, 2001; Nourani et al., *Mol Cell* 21: 7629-40, 2001; Gavin et al., *Nature* 415: 141-7, 2002; Ho et al., *Nature* 415: 180-3, 2002). We did not observe vulval lineage defects after injection of these RNAs into either wild-type or synMuv single mutant backgrounds.

Lastly, bacteria designed to express double-stranded RNA corresponding to the *Gcn5* homolog *Y47G6A.6* (Fraser et al., *Nature* 408: 325-30, 2000) were fed to wild-type and synMuv single mutant hermaphrodites. As described below, we did not observe vulval defects following this treatment.

Deletion allele isolation

Genomic DNA pools from mutagenized worms were screened for deletions essentially as described by (Plasterk et al., *Nat Genet* 17: 119-21, 1997). Deletion mutant animals were isolated from frozen stocks and were
5 backcrossed four times prior to use. *hat-1(n4075)* removes nucleotides +106 to +1115, *epc-1(n4076)* nucleotides +2014 to +2899 and *ssl-1(n4077)* nucleotides +5075 to +5757 of genomic DNA relative to their respective predicted translational start sites.

10 cDNA isolation

We used TITAN ONE-TUBE RT-PCR (Roche Diagnostics, Pleasanton, California) to carry out RT-PCR and recovered *trr-1* and *hat-1* cDNA clones. Existing cDNAs were obtained from the *C. elegans* EST project to determine gene structures of *epc-1*, the *trr-1* 3' end and the *ssl-1* 5' end. We used 5'
15 RACE (5' RACE System v2.0, GIBCO) to determine the 5' ends and SL1 *trans*-spliced leader sequences of *trr-1*, *hat-1*, and *epc-1* transcripts.

Allele sequence

We used PCR-amplified regions of genomic DNA as templates in
20 determining mutant allele sequences. For each allele investigated, we determined the sequences of all exons and splice junctions of the gene in question. All mutations were confirmed by determining the sequence of independently-derived PCR products. All sequences were determined using an automated ABI 373 DNA sequencer (Applied Biosystems).

25

Example III

ssl-1*, a p400 SWI/SNF ATPase homolog, acts redundantly with *lin-15B

TRRAP is a component of the mammalian p400 complex, which contains the p400 SWI/SNF family protein and was identified based on its
30 interaction with the adenovirus E1A oncoprotein (Fuchs et al., *Cell* 106: 297-

307, 2001). Although Tip60 was not present in the purified p400 complex, the Tip60 and p400 complexes share many of the same components and more recent analyses have indicated that p400 and Tip60 can copurify as part of a large p400/Tip60 multisubunit complex (Frank et al., EMBO Rep., 4:575-80, 5 2003).

As discussed in Example II, the *ssl-1* (*ssl*, SWI/SNF-like) gene encodes a homolog of the p400 protein. RNAi of *ssl-1* using standard methods caused fully penetrant embryonic lethality like that observed with *epc-1* (RNAi). zygotic RNAi of *ssl-1*, performed as described above, did not cause defects in 10 vulval development in either class A or class B synMuv backgrounds. In further studies, we isolated a deletion mutation, *n4077*, that removes a portion of the fifth *ssl-1* exon. *ssl-1*(*n4077*) is predicted to encode a truncated protein containing the first 540 amino acids of the 1671 amino acid SSL-1 protein and two unrelated amino acids. *ssl-1*(*n4077*) homozygotes were partially sterile 15 and produced a few inviable embryos, but were not defective in vulval development. *ssl-1*(*n4077*); *lin-15A*(*n767*) mutants were likewise not defective in vulval development, however, *ssl-1*(*n4077*); *lin-15B*(*n744*) mutants often expressed an ectopic vulval cell fate in P8.p. *ssl-1*(*n4077*) likely causes a stronger reduction in gene activity than does *ssl-1* zygotic RNAi, and this 20 stronger reduction unmasks a redundancy between *ssl-1* and *lin-15B*.

***trr-1*; *hat-1*, *trr-1*; *epc-1* and *trr-1*; *ssl-1* double mutants do not show synthetic defects in vulval development**

Whereas synthetic defects in double mutants imply genetic redundancy, 25 the lack of synthetic defects in double mutants can indicate that two genes act in the same genetic pathway. Based on the similar phenotype and genetic interactions of *trr-1*, *hat-1* and *epc-1* mutants and on the copurification of the proteins encoded by their mammalian and yeast counterparts, we hypothesized that *trr-1*, *hat-1* and *epc-1* act together to regulate vulval development. To test 30 this possibility, we constructed double mutants to determine if *hat-1* and *epc-1*

function redundantly with *trr-1*. We measured the numbers of vulval cell fates in *trr-1(n3712); hat-1(n3681)*, *trr-1(n3712); hat-1(n4075)*, and *trr-1(n3712); epc-1(RNAi)* mutants and found that the extent of vulval development observed in these double mutants was similar to that observed in single mutant animals.

- 5 These results suggest that *hat-1* and *epc-1* act in the same genetic pathway as *trr-1*, which by analogy to the class A and class B *lin-35* Rb synMuv pathways, we have named the class C synMuv pathway.

- trr-1; ssl-1* double mutants, and, as described above, *ssl-1; lin-15A* mutants were not synthetically defective in P(3-8).p cell-fate specification. It is
 10 possible that *ssl-1* has both class C and class A synMuv activities, however, additional considerations suggest that *ssl-1* has properties more like those of a class C gene. For instance, *ssl-1; synmuvB* mutants have a defect limited to P8.p, whereas *synmuvA; synmuvB* mutants typically show ectopic vulval cell fates in P3.p, P4.p and P8.p. In addition, *ssl-1* mutants are sterile, and sterility
 15 has not been observed for any class A synMuv gene (Thomas et al., *Development* 126: 3449-59, 1999). These considerations, along with the copurification of the mammalian SSL-1 and HAT-1 counterparts, p400 and Tip60, suggest that *ssl-1* is an atypical class C gene, one that acts redundantly with class B, but not class A synMuv genes.

20

***trr-1, hat-1, epc-1* and *ssl-1* act redundantly with the *lin-35* Rb pathway to antagonize *let-60* Ras signaling**

- Identifying genes involved in cell-fate determination is important for understanding how cells that contain the same genomic information can adopt
 25 different fates during animal development. As they help to distinguish P3.p, P4.p and P8.p from P(5-7).p, *trr-1, hat-1, epc-1* and *ssl-1* are such cell-fate determination genes.

- In many cases, pathways that control cell-fate determination and cell division in invertebrates have been shown to regulate similar processes in
 30 mammals. Pathways that regulate vulval cell-fate specification in *C. elegans*

provide clear examples. A conserved *let-60* Ras pathway induces vulval cell fates, and this pathway is antagonized by an at least partially conserved class B *lin-35* Rb pathway. *trr-1*, *hat-1*, *epc-1* and *ssl-1* act in parallel to *lin-35* Rb and other genes in this pathway to negatively regulate *let-60* Ras signaling. We suggest that the mammalian counterparts of *trr-1*, *hat-1*, *epc-1* and *ssl-1* may similarly act in parallel to Rb and antagonize Ras in the control of cell-fate determination and cell division. It is interesting to note that the p400 complex and Rb-containing complexes are targeted by the adenovirus E1A oncoprotein (Whyte et al., Nature 334:124-9, 1988; Fuchs et al., Cell 106: 297-307, 2001). Our finding regarding *ssl-1* redundancy with a *lin-35* Rb pathway gene suggests that E1A may act in mammals by perturbing the activities of functionally redundant p400 and Rb-containing complexes.

Identification of new class B synMuv genes

On the basis of genetic interactions, the synMuv genes have been grouped into three classes A, B and C. For an animal to show vulval abnormalities, genes representing two of three classes must be dysfunctional. The class B synMuv genes include genes that encode homologs of the mammalian Rb tumor suppressor protein and other proteins that act with Rb in regulating cell-fate specification and division in mammals. We have recently discovered three new class B synMuv genes: *lin(n3628)*, *lin(n4256)*, and *lin-65*. *lin(n3628)* encodes a protein similar to the yeast Set2 histone methyltransferase. The nucleic acid and amino acid sequences of *lin(n3628)* are shown in Figures 23 and 24, respectively. *lin(n4256)* encodes a protein similar to yeast and mammalian SUV39H1 family histone methyltransferases. The nucleic acid and amino acid sequences of *lin(n4256)* are provided in Figures 25 and 26. *lin-65* encodes a protein rich in acidic amino acids. The nucleic acid and amino acid sequences of *lin-65* are provided in Figures 27 and 28.

The striking parallel between the Rb pathway in mammals and the Rb-related pathway we have identified in worms suggests that further characterization of the synthetic Multivulva genes will provide insights into how cell proliferation is regulated in humans. Because synMuv genes encode members of a conserved tumor suppressor pathway that antagonizes a conserved Ras oncogene pathway, the class B synMuv genes are likely to be important in understanding cancer progression in mammals. Provided with the human genome sequence, standard methods can be used to identify mammalian orthologs of newly-identified synMuv genes. Such homologs may act as tumor suppressors or oncogenes in mammals. Genetic enhancer or suppressor screens may be performed to identify new genes which may function in or interface with this Rb-related pathway. Furthermore, using methods described herein, drug screens can be used to identify compounds that affect cell proliferation. Compounds that block the Muv phenotype of synMuv mutant animals are likely to be useful antitumor agents for the treatment of a mammalian neoplasia.

Compounds that stimulate cell division in animals with a single, silent synMuv mutation are likely to be agonists of cell proliferation and may act in a manner analogous to growth factors. Such compounds are useful in the treatment of a subject in need of increased cell proliferation, for example, in a subject that has a disorder characterized by increased cell death, such as Alzheimer's disease, Huntington's disease, stroke, Parkinson's disease, myocardial infarction or congestive heart failure.

Identifying synMuv targets [*Craig: please confirm that this paragraphs reflects our discussion of the screens***]**

The targets of synMuv biological activity, for example, genes that are transcriptionally regulated by a synMuv nucleic acid or polypeptide, are identified using a variety of genetic and molecular approaches. While target identification is discussed below for the class B synMuvs, similar approaches

are used to identify the targets of the class C synMuvs or other transcriptional regulatory systems.

At least two genetic screens can be used to identify class B synMuv targets. Both screens are based on the premise that the class B synMuv proteins negatively regulate transcription. Given that class B synMuv proteins are likely to negatively regulate transcription, one would postulate that the Muv phenotype of synMuv mutants is due to the ectopic expression of class B targets. Loss of function mutations in such targets likely suppress the synMuv phenotype. In one example, a simple F₂ suppression screen is used to identify such targets. In fact, such screens have identified Class B suppressor mutations that may affect such genes. Many of the isolates from these screens are as yet uncharacterized.

In a second example, which would likely identify genes whose expression is negatively regulated by the class B synMuvs, mutagenized class A synMuv F₁ animals are screened for a Muv phenotype. Dominant mutations expected from this screen might affect regulatory sequences bound by synMuv proteins and lead to ectopic expression of the target gene in question. Mutations of this type have been shown to affect the expression of *egl-1*, a gene that promotes programmed cell death in *C. elegans*. These *egl-1(gf)* mutations disrupt a binding site for the TRA-1 transcriptional repressor protein, leading to ectopic *egl-1* expression in the hermaphrodite specific neurons and subsequent programmed cell death (Conradt et al. *Cell* 98:317-27, 1999).

Because transcription factors typically target multiple genes, loss of function of one target may not suppress the phenotype caused by a transcriptional repressor loss of function or, alternatively, recapitulate the phenotype caused by transcriptional activator loss of function. Such challenges are overcome by performing screens in a particularly sensitized genetic background so as to allow the observation of a small effect that may be caused by loss of one target. For example, in one of the screens described above, the Muv phenotype caused by a temperature-sensitive *lin-15AB* allele was

suppressed. A similarly sensitized background may be used for to carry out F₂ suppression and F₁ synMuv screens.

Various molecular approaches involving microarrays are also useful in identifying synMuv targets. In the simplest experiment, expression profiles of synMuv mutants are compared to the wild type. A comparison of synMuv double mutant to the wild type can be problematic because these animals have different amounts of vulval tissue. The generation of vulval tissue likely involves the differential regulation of many genes, only a subset of which might be direct targets of synMuvs. Alternatively, a synMuv single mutant can be compared to a wild-type control. This approach may not succeed if two classes of synMuvs must lose function in order for transcription to be differentially regulated. If mutations in two classes of synMuvs are desired, an appropriate comparison may, for example, be that of a synMuvA; synMuvB; *let-60* Ras triple mutant versus a *let-60* Ras single mutant. These animals would fulfill the requirements of having the same amount of vulval tissue and disabling two classes of synMuvs. Alternatively, chromatin immunoprecipitation (ChIP) combined with microarray analysis may be used. For example, in a preparation of proteins crosslinked to DNA, DPL-1 or EFL-1 could be immunoprecipitated, the crosslink reversed and the resultant DNA amplified and applied to microarrays. Such microarray experiments described above may identify synMuv targets that could be compared to putative *let-60* Ras pathway targets as previously determined by microarray analyses (Romagnolo et al., Dev Biol 247:127-36, 2002). Determining this interface is clearly an important issue as Rb and Ras pathways antagonize each other not only in *C. elegans*, but also during cell cycle progression in cultured mammalian cells (Mittnacht et al., Curr Biol. 7:219-21, 1997; Peeper et al., Nature. 386:177-81, 1997).

Do the synMuv genes act by regulating cell cycle progression?

Many studies of Rb and E2F in mammals have focused on the roles of these proteins in cell cycle regulation. Might the class B synMuv genes, and possibly other classes of synMuv genes regulate vulval development through direct regulation of P(3-8).p cell cycles? While not being tied to a particular theory, the following observations support this possibility. For example, P3.p, P4.p, and P8.p undergo extra cell divisions in synMuv mutants. Additionally, mutations in a subset of class B synMuv genes that includes *dpl-1*, *efl-1*, and *lin-35* Rb have been shown to partially suppress the S phase and cell division defects caused by RNA-mediated interference of the *C. elegans* cyclin D homolog *cyd-1* (Boxem et al., Curr Biol. 12:906-11, 2002). There are other aspects of these observations that complicate a strict cell cycle regulation model. First, not only are there extra P3.p, P4.p and P8.p cell divisions in synMuv mutants, but there are also various changes in the differentiation of P3.p, P4.p and P8.p descendants in synMuv mutants. The synMuv genes therefore appear to regulate a cell fate decision, a component of which is the decision to progress through the cell cycle. Studies of Rb in mammals have indicated that Rb may have a role in halting cell cycle progression and stimulating differentiation during myogenesis (reviewed by Kitzmann Cell Mol Life Sci. 58:571-9, 2001). Second, whereas *dpl-1*, *efl-1*, and *lin-35* Rb mutations can partially suppress defects caused by *cyd-1(RNAi)*, mutations in other class B synMuv genes cannot (Boxem et al., Curr Biol. 12:906-11, 2002). This observation suggests that, if the class B synMuv genes are cell cycle regulators, some of them act in a tissue-specific fashion, for example in P(3-8).p but not in the intestinal cells that were monitored in *cyd-1(RNAi)* studies. Monitoring cell cycle progression in P3.p, P4.p and P8.p will address these issues.

The identification of synMuv transcriptional targets will enable us to identify their mammalian orthologs. Such targets are promising clinical targets for chemotherapeutics for the treatment of neoplasia. In addition, the

identification of synMuv protein-protein interactions is useful in screening for chemotherapeutic drugs that modulate such interactions.

Identification of Additional Mammalian Orthologs

Because the Rb and RAS pathways are conserved between mammals
5 and *C. elegans*, the powerful genetics and genomics of *C. elegans* can be exploited, as described herein, for the systematic identification of mammalian genes that correspond to *C. elegans* genes identified according to methods described herein. Such genes include mammalian orthologs of synMuv class B, and class C genes and their transcriptional targets.

10 Protein sequences corresponding to genes of interest are retrieved from the repositories of *C. elegans* sequence information at the wormbase web site. The *C. elegans* protein or nucleic acid sequence is then used for standard [BLASTP] or [tblastn] searching using the NCBI website. The protein sequence corresponding to the top mammalian candidate produced by tblastn is
15 retrieved from Genbank and is used for BLASTp search of *C. elegans* proteins using the wormbase website. These methods allow us to identify mammalian orthologs of worm genes revealed by our genetic analysis.

An ortholog is a protein that is functionally related to a reference sequence. Such orthologs might be expected to functionally substitute for one
20 another. For example, expression of a mammalian ortholog of a *C. elegans* gene, when expressed in a worm having a mutation in the *C. elegans* gene, might be expected to partially or completely rescue the worm phenotype.

RNAi in mammalian cell lines

25 RNAi has been used extensively to deplete mRNAs in mammalian cell culture (Elbashir et al., Nature 411:494-8, 2001). Mammalian orthologs of class C synMuv genes can be identified using RNAi, for example, in mammalian cultured cells. Briefly, an inhibitory nucleic acid is introduced into a mammalian cell having a mutation in a class A or class B synMuv gene, for
30 example, by lipofection. Such cells are then assayed for increased levels of cell

proliferation relative to control cells not contacted with an inhibitory nucleic acid. An increased level of proliferation in mammalian cells contacted with the inhibitory nucleic acid identifies the corresponding target gene as a class C synMuv gene.

5

Microarrays

The class B and class C genes described herein, are useful in identifying their transcriptional regulatory targets. Such targets may be identified using microarrays in combination with chromatin immunoprecipitation (chIP) as described herein. Such methods are described in U.S. Patent 6,503,717, 6,410,243, and 6,610,489, hereby incorporated by reference. A nucleic acid target of a class B or class C synMuv polypeptide will likely have a mammalian ortholog. Such an ortholog represents a promising target for the development of novel chemotherapeutics for the treatment of a neoplasia.

15 The array elements, which are preferably derived from the *C. elegans* genome, are organized in an ordered fashion such that each element is present at a specified location on the substrate. Useful substrate materials include membranes, composed of paper, nylon or other materials, filters, chips, glass slides, and other solid supports. The ordered arrangement of the array elements allows hybridization patterns and intensities to be interpreted as expression levels of particular genes or proteins. Methods for making nucleic acid microarrays are known to the skilled artisan and are described, for example, in U.S. Patent No. 5,837,832, Lockhart, et al. (Nat. Biotech. 14:1675-1680, 1996), and Schena, et al. (Proc. Natl. Acad. Sci. 93:10614-10619, 1996), herein incorporated by reference. Methods for making polypeptide microarrays are described, for example, by Ge (Nucleic Acids Res. 28:e3.i-e3.vii, 2000), MacBeath et al., (Science 289:1760-1763, 2000), Zhu et al. (Nature Genet. 26:283-289), and in U.S. Patent No. 6,436,665, hereby incorporated by reference.

30

Nucleic acid microarrays

To produce a nucleic acid microarray oligonucleotides may be synthesized or bound to the surface of a substrate using a chemical coupling procedure and an ink jet application apparatus, as described in PCT application
5 W095/251116 (Baldeschweiler et al.), incorporated herein by reference.

Alternatively, a gridded array may be used to arrange and link cDNA fragments or oligonucleotides to the surface of a substrate using a vacuum system, thermal, UV, mechanical or chemical bonding procedure.

A nucleic acid molecule (e.g. RNA or DNA) derived from a biological
10 sample, such as a cultured cell, a tissue specimen, or other source, may be used to produce a hybridization probe as described herein. The mRNA is isolated according to standard methods, and cDNA is produced and used as a template to make complementary RNA suitable for hybridization using standard methods. The RNA is amplified in the presence of fluorescent nucleotides, and
15 the labeled probes are then incubated with the microarray to allow the probe sequence to hybridize to complementary oligonucleotides bound to the microarray.

Incubation conditions are adjusted such that hybridization occurs with precise complementary matches or with various degrees of less
20 complementarity depending on the degree of stringency employed. For example, stringent salt concentration will ordinarily be less than about 750 mM NaCl and 75 mM trisodium citrate, preferably less than about 500 mM NaCl and 50 mM trisodium citrate, and most preferably less than about 250 mM NaCl and 25 mM trisodium citrate. Low stringency hybridization can be
25 obtained in the absence of organic solvent, e.g., formamide, while high stringency hybridization can be obtained in the presence of at least about 35% formamide, and most preferably at least about 50% formamide. Stringent temperature conditions will ordinarily include temperatures of at least about 30°C, more preferably of at least about 37°C, and most preferably of at least
30 about 42°C. Varying additional parameters, such as hybridization time, the

concentration of detergent, e.g., sodium dodecyl sulfate (SDS), and the inclusion or exclusion of carrier DNA, are well known to those skilled in the art. Various levels of stringency are accomplished by combining these various conditions as needed. In a preferred embodiment, hybridization will occur at
5 30°C in 750 mM NaCl, 75 mM trisodium citrate, and 1% SDS. In a more preferred embodiment, hybridization will occur at 37°C in 500 mM NaCl, 50 mM trisodium citrate, 1% SDS, 35% formamide, and 100 µg/ml denatured salmon sperm DNA (ssDNA). In a most preferred embodiment, hybridization will occur at 42°C in 250 mM NaCl, 25 mM trisodium citrate, 1% SDS, 50%
10 formamide, and 200 µg/ml ssDNA. Useful variations on these conditions will be readily apparent to those skilled in the art.

The removal of nonhybridized probes may be accomplished, for example, by washing. The washing steps that follow hybridization can also vary in stringency. Wash stringency conditions can be defined by salt
15 concentration and by temperature. As above, wash stringency can be increased by decreasing salt concentration or by increasing temperature. For example, stringent salt concentration for the wash steps will preferably be less than about 30 mM NaCl and 3 mM trisodium citrate, and most preferably less than about 15 mM NaCl and 1.5 mM trisodium citrate. Stringent temperature conditions
20 for the wash steps will ordinarily include a temperature of at least about 25°C, more preferably of at least about 42°C, and most preferably of at least about 68°C. In a preferred embodiment, wash steps will occur at 25°C in 30 mM NaCl, 3 mM trisodium citrate, and 0.1% SDS. In a more preferred embodiment, wash steps will occur at 42°C in 15 mM NaCl, 1.5 mM trisodium
25 citrate, and 0.1% SDS. In a most preferred embodiment, wash steps will occur at 68°C in 15 mM NaCl, 1.5 mM trisodium citrate, and 0.1% SDS. Additional variations on these conditions will be readily apparent to those skilled in the art.

A detection system may be used to measure the absence, presence, and
30 amount of hybridization for all of the distinct sequences simultaneously (e.g.,

Heller et al., Proc. Natl. Acad. Sci. 94:2150-2155, 1997). Preferably, a scanner is used to determine the levels and patterns of fluorescence.

Protein Microarrays

5 Families of proteins, such as those encoded by the genes described herein, or their orthologs, may be analyzed using protein microarrays. Such arrays are useful in high-throughput low-cost screens to identify peptide or candidate compounds that bind a polypeptide of the invention, or fragment thereof. Typically, protein microarrays feature a protein, or fragment thereof,
10 bound to a solid support. Suitable solid supports include membranes (e.g., membranes composed of nitrocellulose, paper, or other material), polymer-based films (e.g., polystyrene), beads, or glass slides. For some applications, proteins (e.g., polypeptides encoded by class B or class C synMuv gene or antibodies against such polypeptides) are spotted on a substrate using any
15 convenient method known to the skilled artisan (e.g., by hand or by inkjet printer). Preferably, such methods retain the biological activity or function of the protein bound to the substrate

The protein microarray is hybridized with a detectable probe. Such probes can be polypeptide, nucleic acid, or small molecules. For some
20 applications, polypeptide and nucleic acid probes are derived from a biological sample taken from a patient, such as a homogenized tissue sample (e.g. a tissue sample obtained by biopsy); or cultured cells (e.g., lymphocytes). Probes can also include antibodies, candidate peptides, nucleic acids, or small molecule compounds derived from a peptide, nucleic acid, or chemical library.
25 Hybridization conditions (e.g., temperature, pH, protein concentration, and ionic strength) are optimized to promote specific interactions. Such conditions are known to the skilled artisan and are described, for example, in Harlow, E. and Lane, D., Using Antibodies : A Laboratory Manual. 1998, New York: Cold Spring Harbor Laboratories. After removal of non-specific probes, specifically
30 bound probes are detected, for example, by fluorescence, enzyme activity (e.g.,

an enzyme-linked colorimetric assay), direct immunoassay, radiometric assay, or any other suitable detectable method known to the skilled artisan.

Screening Assays

5 As discussed above, *C. elegans* class B and class C synMuv genes and their encoded proteins function in chromatin remodeling and antagonize the RAS pathway. Given that mechanisms for controlling mammalian cell cycle regulation and *C. elegans* vulval development are highly conserved, *C. elegans* and components of the *C. elegans* synMuv pathway are useful in screening
10 methods for chemotherapeutics and for the identification of novel clinical targets.

 Compounds that modulate the function of a Class B, or Class C synMuv nucleic acid or of their encoded proteins are likely to be useful in treating neoplasias. Based on this discovery, screening assays may be carried out to
15 identify compounds that modulate the action of a polypeptide or the expression of a nucleic acid sequence of the invention. Such compounds are useful in treating a neoplasia. The method of screening may involve high-throughput techniques. In addition, these screening techniques may be carried out in cultured mammalian cells or in animals (e.g., nematodes).

20 Any number of methods are available for carrying out such screening assays. In one working example, candidate compounds are added at varying concentrations to the culture medium of cultured cells expressing one of the nucleic acid sequences described herein. Gene expression is then measured, for example, by standard Northern blot analysis (Ausubel et al., supra) or RT-
25 PCR, using any appropriate fragment prepared from the nucleic acid molecule as a hybridization probe. The level of gene expression in the presence of the candidate compound is compared to the level measured in a control culture medium lacking the candidate molecule. A compound that promotes a decrease in the expression of a nucleic acid sequence disclosed herein or a
30 functional equivalent is considered useful in the invention; such a molecule

may be used, for example, as a therapeutic to delay or ameliorate human diseases associated with neoplasia or inappropriate cell cycle regulation. Such cultured cells include nematode cells (for example, *C. elegans* cells), mammalian, or insect cells.

5 In another working example, the effect of candidate compounds may be measured at the level of polypeptide production using the same general approach and standard immunological techniques, such as Western blotting or immunoprecipitation with an antibody specific for a polypeptide of the invention. For example, immunoassays may be used to detect or monitor the
10 expression of at least one of the polypeptides of the invention in an organism. Polyclonal or monoclonal antibodies (produced by standard techniques) that are capable of binding to such a polypeptide may be used in any standard immunoassay format (e.g., ELISA, Western blot, or RIA assay) to measure the level of the polypeptide. A compound that promotes a decrease in the
15 expression of the polypeptide is considered particularly useful. Again, such a molecule may be used, for example, as a therapeutic to ameliorate neoplasia.

In one example, candidate compounds are screened for those that specifically bind to and antagonize a synMuv B or synMuv C polypeptide. Such an interaction can be readily assayed using any number of standard
20 binding techniques and functional assays (e.g., those described in Ausubel et al., supra). For example, a candidate compound may be tested *in vitro* for interaction and binding with a polypeptide of the invention and its ability to modulate the cell cycle or decrease cell proliferation may be assayed by any standard technique (e.g., a *C. elegans* synMuv assay).

25 In one particular working example, a candidate compound that binds to a polypeptide may be identified using a chromatography-based technique. For example, a recombinant polypeptide of the invention may be purified by standard techniques from cells engineered to express the polypeptide (e.g., those described above) and may be immobilized on a column. A solution of
30 candidate compounds is then passed through the column, and a compound

specific for the polypeptide is identified on the basis of its ability to bind to the polypeptide and be immobilized on the column. To isolate the compound, the column is washed to remove non-specifically bound molecules, and the compound of interest is then released from the column and collected.

- 5 Compounds isolated by this method (or any other appropriate method) may, if desired, be further purified (e.g., by high performance liquid chromatography). In addition, these candidate compounds may be tested for their ability to cause cell death using any assay known to the skilled artisan. Compounds isolated by this approach may also be used, for example, as therapeutics to delay or
- 10 ameliorate human diseases associated with neoplasia. Compounds that are identified as binding to polypeptides of the invention with an affinity constant less than or equal to 10 mM are considered particularly useful in the invention.

- Potential antagonists include organic molecules, peptides, peptide mimetics, polypeptides, nucleic acids, and antibodies that bind to a nucleic acid
- 15 sequence or polypeptide of the invention and thereby increase or decrease its activity. Potential antagonists also include small molecules that bind to and occupy the binding site of the polypeptide thereby preventing binding to cellular binding molecules, such that normal biological activity is prevented.

- Each of the DNA sequences provided herein may also be used in the
- 20 discovery and development of therapeutic lead compounds. The encoded protein, upon expression, can be used as a target for the screening of therapeutics for the treatment of neoplasia. Additionally, the DNA sequences encoding the amino terminal regions of the encoded protein or Shine-Delgarno or other translation facilitating sequences of the respective mRNA can be used
- 25 to construct antisense, dsRNAs, or siRNA sequences to control the expression of the coding sequence of interest. Such sequences may be isolated by standard techniques (Ausubel et al., *supra*). The antagonists of the invention may be employed, for instance, to delay or ameliorate human diseases associated with neoplasia.

Optionally, compounds identified in any of the above-described assays may be confirmed as useful in delaying or ameliorating human diseases associated with neoplasia or inappropriate cell cycle regulation in either standard tissue culture methods or animal models and, if successful, may be
5 used as therapeutics for the treatment of neoplasia.

Small molecules of the invention preferably have a molecular weight below 2,000 daltons, more preferably between 300 and 1,000 daltons, and most preferably between 400 and 700 daltons. It is preferred that these small molecules are organic molecules.

10

Test Compounds and Extracts

In general, compounds capable of delaying or ameliorating human diseases associated with neoplasia are identified from large libraries of both natural product or synthetic (or semi-synthetic) extracts or chemical libraries
15 according to methods known in the art. Those skilled in the field of drug discovery and development will understand that the precise source of test extracts or compounds is not critical to the screening procedure(s) of the invention. Compounds used in screens may include known compounds (for example, known therapeutics used for other diseases or disorders).

20 Alternatively, virtually any number of unknown chemical extracts or compounds can be screened using the methods described herein. Examples of such extracts or compounds include, but are not limited to, plant-, fungal-, prokaryotic- or animal-based extracts, fermentation broths, and synthetic compounds, as well as modification of existing compounds. Numerous
25 methods are also available for generating random or directed synthesis (e.g., semi-synthesis or total synthesis) of any number of chemical compounds, including, but not limited to, saccharide-, lipid-, peptide-, and nucleic acid-based compounds. Synthetic compound libraries are commercially available from Brandon Associates (Merrimack, NH) and Aldrich Chemical (Milwaukee,
30 WI). Alternatively, libraries of natural compounds in the form of bacterial,

5 fungal, plant, and animal extracts are commercially available from a number of sources, including Biotics (Sussex, UK), Xenova (Slough, UK), Harbor Branch Oceanographics Institute (Ft. Pierce, FL), and PharmaMar, U.S.A. (Cambridge, MA). In addition, natural and synthetically produced libraries are produced, if desired, according to methods known in the art, e.g., by standard extraction and fractionation methods. Furthermore, if desired, any library or compound is readily modified using standard chemical, physical, or biochemical methods.

10 In addition, those skilled in the art of drug discovery and development readily understand that methods for dereplication (e.g., taxonomic dereplication, biological dereplication, and chemical dereplication, or any combination thereof) or the elimination of replicates or repeats of materials already known to function in neoplasia should be employed whenever possible.

15 When a crude extract is found to decrease cell proliferation or to suppress a synMuv phenotype, further fractionation of the positive lead extract is necessary to isolate chemical constituents responsible for the observed effect. Thus, the goal of the extraction, fractionation, and purification process is the careful characterization and identification of a chemical entity within the crude extract that inhibits cell proliferation or suppresses a synMuv phenotype. Methods of fractionation and purification of such heterogenous extracts are known in the art. If desired, compounds shown to be useful agents to delay or ameliorate human diseases associated with neoplasia are chemically modified according to methods known in the art.

Pharmaceutical Therapeutics

25 The invention provides a simple means for identifying compositions (including nucleic acids, peptides, small molecule inhibitors, and mimetics) capable of acting as therapeutics for the treatment of a neoplastic disease. Accordingly, a chemical entity discovered to have medicinal value using the methods described herein is useful as a drug or as information for structural modification of existing compounds, e.g., by rational drug design. Such

30

methods are useful for screening compounds having an effect on a variety of diseases characterized by inappropriate cell cycle regulation.

For therapeutic uses, the compositions or agents identified using the methods disclosed herein may be administered systemically, for example, formulated in a pharmaceutically-acceptable buffer such as physiological saline. Preferable routes of administration include, for example, subcutaneous, intravenous, interperitoneally, intramuscular, or intradermal injections that provide continuous, sustained levels of the drug in the patient. Treatment of human patients or other animals will be carried out using a therapeutically effective amount of a neoplastic disease therapeutic in a physiologically-acceptable carrier. Suitable carriers and their formulation are described, for example, in Remington's Pharmaceutical Sciences by E.W. Martin. The amount of the therapeutic agent to be administered varies depending upon the manner of administration, the age and body weight of the patient, and with the clinical symptoms of the neoplastic disease. Generally, amounts will be in the range of those used for other agents used in the treatment of a neoplastic disease, although in certain instances lower amounts will be needed because of the increased specificity of the compound. A compound is administered at a dosage that controls the clinical or physiological symptoms of a neoplastic disease as determined by, for example, measuring tumor size, cell proliferation, or metastasis.

Formulation of Pharmaceutical Compositions

Administration of a compound may be by any suitable means that is effective for the treatment of a neoplastic disease. Generally, compounds are admixed with a suitable carrier substance, and are generally present in an amount of 1-95% by weight of the total weight of the composition. The composition may be provided in a dosage form that is suitable for oral, parenteral (e.g., intravenous, intramuscular, subcutaneous), rectal, transdermal, nasal, vaginal, inhalant, or ocular administration. Thus, the composition may

be in form of, e.g., tablets, capsules, pills, powders, granulates, suspensions, emulsions, solutions, gels including hydrogels, pastes, ointments, creams, plasters, drenches, delivery devices, suppositories, enemas, injectables, implants, sprays, or aerosols. The pharmaceutical compositions may be formulated according to conventional pharmaceutical practice (see, e.g., Remington: The Science and Practice of Pharmacy, (20th ed.) ed. A.R. Gennaro, 2000, Lippincott Williams & Wilkins, Philadelphia, PA. and Encyclopedia of Pharmaceutical Technology, eds. J. Swarbrick and J. C. Boylan, 1988-2002, Marcel Dekker, New York).

10

Other Embodiments

From the foregoing description, it will be apparent that variations and modifications may be made to the invention described herein to adapt it to various usages and conditions. Such embodiments are also within the scope of the following claims.

15

All publications mentioned in this specification are herein incorporated by reference to the same extent as if each independent publication was specifically and individually indicated to be incorporated by reference.

20

What is claimed is:

Claims

1. A method for identifying a compound that treats a neoplasia, said method comprising the steps of:

5 (a) contacting a cell comprising a mutation in a Class B synMuv gene selected from the group consisting of: *mep-1*, *lin(n3628)*, *lin(n4256)*, and *lin-65* and a second mutation in a synthetic multivulval gene, or an ortholog thereof, with a candidate compound;

(b) detecting a phenotypic alteration in said contacted cell relative to a control cell; wherein a candidate compound that alters the phenotype of said
10 contacted cell relative to said control cell is a compound that treats a neoplasia.

2. The method of claim 1, wherein said cell is in a nematode.

3. The method of claim 2, wherein said phenotypic alteration is an
15 alteration in a multivulval phenotype.

4. The method of claim 2, wherein said phenotypic alteration is an alteration in sterility.

20 5. The method of claim 1, wherein said synthetic multivulval gene is a synMuv class A gene.

6. The method of claim 1, wherein said cell is an isolated mammalian cell.
25

7. The method of claim 1, wherein said phenotypic alteration is a decrease in cell proliferation.

8. A method for identifying a candidate compound that treats a neoplasia, said method comprising:

- (a) providing a cell having a mutation in a Class B synMuv gene selected from the group consisting of *mep-1*, *lin(n3628)*, *lin(n4256)*, and *lin-65*
5 and having a second mutation in a synMuv nucleic acid or ortholog thereof;
(b) contacting said cell with a candidate compound; and
(c) detecting a decrease in proliferation of said cell contacted with said candidate compound relative to a control cell not contacted with said candidate compound, wherein a decrease in proliferation identifies said candidate
10 compound as a candidate compound that treats a neoplasia.

9. The method of claim 8, wherein said cell is in a nematode.

10. The method of claim 9, wherein said decrease in proliferation is
15 detected by detecting inhibition of a Muv phenotype.

11. The method of claim 8, wherein said cell has a mutation in Dp, E2F, or histone deacetylase.

20 12. The method of claim 8, wherein said cell is an isolated mammalian cell.

13. A method of identifying a compound that treats a neoplasia, said method comprising:

(a) providing a cell expressing a nucleic acid having at least 95% identity to a Class B synMuv gene selected from the group consisting of: *mep-1*, *lin(n3628)*, *lin(n4256)*, and *lin-65*;

(b) contacting said cell with a candidate compound; and

(c) monitoring the expression of said nucleic acid, an alteration in the level of expression of said nucleic acid indicates that said candidate compound is a compound that treats a neoplasia.

10

14. The method of claim 13, wherein said gene comprises a reporter gene.

15. The method of claim 13, wherein said reporter gene comprises *lacZ*, *gfp*, CAT, or luciferase.

16. The method of claim 13, wherein said expression is monitored by assaying protein level.

17. The method of claim 13, wherein said expression is monitored by assaying nucleic acid level.

18. The method of claim 13, wherein said cell is in a nematode.

25

19. A method for identifying a candidate compound that treats a neoplasia, said method comprising:

(a) providing a cell expressing a Class B synMuv gene selected from the group consisting of: *mep-1*, *lin(n3628)*, *lin(n4256)*, and *lin-65*;

5 (b) contacting said cell with a candidate compound; and

(c) comparing the expression of said polypeptide in said cell contacted with said candidate compound to a control cell not contacted with said candidate compound, wherein an increase in the expression of said polypeptide identifies said candidate compound as a candidate compound that treats a
10 neoplasia..

20. The method of claim 19, wherein said cell is in a nematode.

21. The method of claim 19, wherein said expression is monitored
15 with an immunological assay.

22. A method for identifying a candidate compound that treats a neoplasia, said method comprising:

(a) providing a cell expressing a Class B synMuv polypeptide selected
20 from the group consisting of: MEP-1, LIN(n3628), LIN(n4256), and LIN-65, said method comprising;

(b) contacting said cell with a candidate compound; and

(c) comparing the biological activity of said polypeptide in said cell contacted with said candidate compound to a control cell not contacted with
25 said candidate compound, wherein an increase in the biological activity of said polypeptide identifies said candidate compound as a candidate compound that treats a neoplasia.

23. The method of claim 22, wherein said biological activity is
30 monitored with an enzymatic assay.

24. The method of claim 22, wherein said biological activity is monitored with an immunological assay.

5 25. The method of claim 22, wherein said biological activity is monitored with a nematode bioassay.

26. A method of identifying a nucleic acid target of class B synMuv biological activity, said method comprising:

- 10 (a) mutagenizing a *C. elegans* comprising mutations in a Class B synMuv gene selected from the group consisting of: *mep-1*, *lin(n3628)*, *lin(n4256)*, and *lin-65* and in a Class A synMuv gene;
- (b) allowing said *C. elegans* to reproduce; and
- (c) selecting a *C. elegans* comprising a mutation that suppresses a synMuv phenotype; wherein said mutation identifies a nucleic acid target of class B synMuv biological activity.
- 15

27. A method of identifying a nucleic acid target of class B synMuv biological activity, said method comprising:

- 20 (a) providing a microarray comprising fragments of nematode nucleic acids;
- (b) contacting said microarray with detectably labeled nucleic acids derived from a nematode comprising a mutation in a Class B synMuv gene selected from the group consisting of: *mep-1*, *lin(n3628)*, *lin(n4256)*, and *lin-65* gene;
- 25 (c) detecting an alteration in the expression of at least one nucleic acid of a *C. elegans* comprising a mutation in said Class B synMuv gene relative to the expression of said nucleic acid in a control nematode, wherein an alteration in said expression identifies said nucleic acid as a nucleic acid target of class B synMuv biological activity.
- 30

28. The method of claim 27, wherein said *C. elegans* further comprises a mutation in a second synMuv gene.

5 29. The method of claim 27, wherein said *C. elegans* further comprises a mutation in a gene that results in a Vulvaless (Vul) phenotype.

30. A method for identifying a nucleic acid that binds a synMuv class B polypeptide, said method comprising:

- 10 (a) providing nucleic acids derived from a nematode cell;
- (b) crosslinking said nucleic acids and their associated proteins to form a nucleic acid-protein complex;
- (c) contacting said nucleic acid-protein complex with an antibody against a polypeptide selected from the group consisting of MEP-1,
- 15 LIN(n3628), LIN(n4256), and LIN-65;
- (d) purifying said nucleic acid-protein complex using an immunological method; and
- (e) isolating said nucleic acid, wherein said isolated nucleic acid is a nucleic acid that binds a synMuv class B polypeptide.

20

31. The method of claim 30, further comprising the following steps:

- (f) detectably labeling the nucleic acid of step (e);
- (g) contacting a microarray comprising *C. elegans* nucleic acid fragments with said detectably labeled nucleic acid; and
- 25 (h) detecting binding of said detectably labeled nucleic acid, wherein said binding identifies said nucleic acid as a nucleic acid that binds a synMuv class B polypeptide.

32. A vector comprising a nucleic acid having at least 95% identity to a Class B synMuv gene selected from the group consisting of: *mep-1*, *lin(n3628)*, *lin(n4256)*, and *lin-65*.

5 33. The vector of claim 32, wherein said synMuv gene is *mep-1* (SEQ ID NO:2).

34. The nucleic acid of claim 33, wherein said synMuv gene comprises a mutation selected from the group consisting of *n3680*, *n3702*, and
10 *n3703*.

35. The vector of claim 32, wherein said synMuv gene is *lin(n3628)* (SEQ ID NO:24).

15 36. The vector of claim 32, wherein said synMuv gene is *lin(n4256)* (SEQ ID NO:26).

37. The vector of claim 36, wherein said synMuv gene is *lin-65* (SEQ ID NO:28).

20

38. An isolated cell comprising the vector of claim 32.

39. A nematode comprising the nucleic acid of claim 32.

25 40. A nematode comprising a mutation in a Class B synMuv gene selected from the group consisting of: *mep-1*, *lin(n3628)*, *lin(n4256)*, and *lin-65*.

41. The nematode of claim 40, wherein said mutation is a *mep-1*
30 mutation selected from the group consisting of *n3680*, *n3702*, and *n3703*.

42. A purified nucleic acid comprising a sequence that hybridizes under high stringency conditions to a Class B synMuv nucleic acid selected from the group consisting of: *mep-1*, *lin(n3628)*, *lin(n4256)*, and *lin-65*.

5

38. An antibody against a Class B synMuv polypeptide selected from the group consisting of: MEP-1, LIN(n3628), LIN(n4256), and LIN-65.

38. A method for identifying a compound that treats a condition
10 characterized by inappropriate cell death, said method comprising the steps of:

(a) contacting a nematode comprising a mutation in a Class B synMuv gene selected from the group consisting of: *mep-1*, *lin(n3628)*, *lin(n4256)*, and *lin-65* with a candidate compound;

(b) detecting a *muv* phenotype in said contacted nematode relative to a
15 control nematode; wherein a candidate compound that alters the phenotype of said contacted nematode relative to said control nematode is a compound that treats a condition characterized by inappropriate cell death.

39. The method of claim 38, wherein said cell is in a nematode.

20

40. The method of claim 38, wherein said alteration is an alteration in synMuv phenotype.

41. A method for identifying a compound that treats a neoplasia, said method comprising the steps of:

- (a) contacting a cell comprising a mutation in a gene encoding KIAA1732 and a second mutation in a synMuv nucleic acid, or an ortholog thereof, with a candidate compound;
- (b) detecting a phenotypic alteration in said contacted cell relative to a control cell; wherein a candidate compound that alters the phenotype of said contacted cell relative to said control cell is a compound that treats a neoplasia.

42. The method of claim 1, wherein said synthetic multivulval gene is a synMuv class A gene.

43. The method of claim 1, wherein said cell is an isolated mammalian cell.

44. The method of claim 1, wherein said phenotypic alteration is a decrease in cell proliferation.

45. A method for identifying a candidate compound that treats a neoplasia, said method comprising:

- (a) providing a cell having a mutation in a nucleic acid encoding KIAA1732 and having a second mutation in a synMuv nucleic acid, or ortholog thereof;
- (b) contacting said cell with a candidate compound; and
- (c) detecting a decrease in proliferation of said cell contacted with said candidate compound relative to a control cell not contacted with said candidate compound, wherein a decrease in proliferation identifies said candidate compound as a candidate compound that treats a neoplasia.

46. The method of claim 8, wherein said cell has a mutation in Dp, E2F, or histone deacetylase.

47. The method of claim 5, wherein said cell is an isolated
5 mammalian cell.

48. A method of identifying a compound that treats a neoplasia, said method comprising:

- 10 (a) providing a cell expressing a nucleic acid having at least 95% identity to a nucleic acid that encodes KIAA1732; and
- (b) contacting said cell with a candidate compound; and
- (c) monitoring the expression of said nucleic acid, an alteration in the level of expression of said nucleic acid indicates that said candidate compound is a compound that treats a neoplasia.

15

49. The method of claim 8, wherein said gene comprises a reporter gene.

50. The method of claim 8, wherein said reporter gene comprises *lacZ*,
20 *gfp*, CAT, or luciferase.

51. The method of claim 8, wherein said expression is monitored by assaying protein level.

25 52. The method of claim 8, wherein said expression is monitored by assaying nucleic acid level.

53. The method of claim 12, wherein said cell is an isolated mammalian cell.

30

54. A method for identifying a candidate compound that treats a neoplasia, said method comprising:

- (a) providing a cell expressing a KIAA1732 polypeptide;
- (b) contacting said cell with a candidate compound; and
- 5 (c) comparing the expression of said polypeptide in said cell contacted with said candidate compound to a control cell not contacted with said candidate compound, wherein an increase in the expression of said polypeptide identifies said candidate compound as a candidate compound that treats a neoplasia.

10

55. The method of claim 54, wherein said cell is an isolated mammalian cell.

56. The method of claim 54, wherein said expression is monitored
15 with an immunological assay.

57. A method for identifying a candidate compound that treats a neoplasia, said method comprising:

- (a) providing a cell expressing a KIAA1732 polypeptide;
- 20 (b) contacting said cell with a candidate compound; and
- (c) comparing the biological activity of said polypeptide in said cell contacted with said candidate compound to a control cell not contacted with said candidate compound, wherein an increase in the biological activity of said polypeptide identifies said candidate compound as a candidate compound that
25 treats a neoplasia.

58. The method of claim 57, wherein said biological activity is monitored with an enzymatic assay.

59. The method of claim 57, wherein said biological activity is monitored with an immunological assay.

60. The method of claim 57, wherein said biological activity is
5 methyl transferase activity.

61. A method for identifying a nucleic acid that binds KIAA1732, said method comprising:

- (a) providing nucleic acids derived from a mammalian cell;
- 10 (b) crosslinking said nucleic acids and their associated proteins to form a nucleic acid-protein complex;
- (c) contacting said nucleic acid-protein complex with an anti-KIAA1732 antibody;
- (d) purifying said nucleic acid-protein complex using an immunological
15 method; and
- (e) isolating said nucleic acid, wherein said isolated nucleic acid is a nucleic acid that binds KIAA1732.

62. The method of claim 61, further comprising the following steps:
20 (f) detectably labeling the nucleic acid of step (e);
(g) contacting a microarray comprising human nucleic acid fragments with said detectably labeled nucleic acid; and
(h) detecting binding of said detectably labeled nucleic acid, wherein said binding identifies said nucleic acid as a nucleic acid that binds KIAA1732.

25

66. A vector comprising a nucleic acid having at least 95% identity to (SEQ ID NO:30).

67. An isolated cell comprising the vector of claim 26.
30

68. A method for identifying a compound that treats a neoplasia, said method comprising the steps of:

(a) contacting a nematode comprising a mutation in a Class C synMuv gene selected from the group consisting of *trr-1*, *hat-1*, *epc-1*, and *ssl-1* with a candidate compound; and

(b) detecting an altered phenotype in said contacted nematode relative to a control nematode; wherein a candidate compound that alters the phenotype of said contacted nematode relative to said control nematode is a compound that treats a neoplasia.

10

69. The method of claim 68, wherein said alteration is an alteration in vulval phenotype.

70. The method of claim 68, wherein said alteration is an alteration in sterility.

15

71. The method of claim 68, wherein said synMuv class C gene is *trr-1*.

72. The method of claim 71, wherein said mutations are selected from the group consisting of *n3630*, *n3637*, *n3704*, *n3708*, *n3709*, and *n3712*.

20

73. A method for identifying a candidate compound that treats a neoplasia, said method comprising:

(a) providing a cell having a mutation in a Class C synMuv gene selected from the group consisting of *trr-1*, *hat-1*, *epc-1*, and *ssl-1* nucleic acid and having a second mutation in a synMuv nucleic acid or ortholog thereof;

25

(b) contacting said cell with a candidate compound; and

(c) detecting a decreased proliferation of said cell contacted with said candidate compound relative to a control cell not contacted with said candidate

30

compound, wherein a decrease in proliferation identifies said candidate compound as a candidate compound that treats a neoplasia.

74. The method of claim 73, wherein said cell is in a nematode.

5

75. The method of claim 73, wherein said nematode displays an alteration in a synMuv phenotype.

76. The method of claim 73, wherein said cell comprises a mutation
10 in a class A or class B synMuv gene.

77. A method for identifying a compound that treats a neoplasia, said method comprising the steps of:

(a) contacting a nematode comprising a mutation in a Class C synMuv
15 gene selected from the group consisting of *trr-1*, *hat-1*, *epc-1*, and *ssl-1* and a second mutation in a Class A synthetic multivulval gene with a candidate compound;

(b) detecting an altered phenotype in said contacted nematode relative to a control nematode; wherein a candidate compound that alters the phenotype
20 of said contacted nematode relative to said control nematode is a compound that treats a neoplasia.

78. The method of claim 77, wherein said alteration is an alteration in synMuv phenotype.

25

79. The method of claim 77, wherein said alteration is an alteration in sterility.

80. A method for identifying a compound that treats a neoplasia, said method comprising the steps of:

(a) contacting a nematode comprising a mutation in a Class C synMuv gene selected from the group consisting of *trr-1*, *hat-1*, *epc-1*, and *ssl-1* and a
5 second mutation in a Class B synthetic multivulval gene with a candidate compound;

(b) detecting an altered phenotype in said contacted nematode relative to a control nematode; wherein a candidate compound that alters the phenotype of said contacted nematode relative to said control nematode is a compound
10 that treats a neoplasia.

81. The method of claim 80, wherein said alteration is an alteration in synMuv phenotype.

15 82. The method of claim 80, wherein said alteration is an alteration in sterility.

83. A method for identifying a candidate compound that treats a neoplasia, said method comprising:

20 (a) providing a cell having a mutation in a Class C synMuv gene selected from the group consisting of *trr-1*, *hat-1*, *epc-1*, and *ssl-1* and having a second mutation in a synMuv gene or ortholog thereof;

(b) contacting said cell with a candidate compound; and

(c) detecting a decreased proliferation of said cell contacted with said
25 candidate compound relative to a control cell not contacted with said candidate compound, wherein a decrease in proliferation identifies said candidate compound as a candidate compound that treats a neoplasia.

84. The method of claim 83, wherein said cell is in a nematode.

30

85. The method of claim 83, wherein said nematode displays an alteration in a synMuv phenotype.

86. A method of identifying a compound that treats a neoplasia, said
5 method comprising:

(a) providing a cell expressing a nucleic acid having at least 95% identity to a Class C synMuv nucleic acid selected from the group consisting of *trr-1*, *hat-1*, *epc-1*, and *ssl-1*;

(b) contacting said cell with a candidate compound; and

10 (c) monitoring the expression of said nucleic acid, an alteration in the level of expression of said nucleic acid indicates that said candidate compound is a compound that treats a neoplasia.

87. The method of claim 86, wherein said gene comprises a reporter
15 gene.

88. The method of claim 86, wherein said reporter gene comprises *lacZ*, *gfp*, CAT, or luciferase.

20 89. The method of claim 86, wherein said expression is monitored by assaying protein level.

90. The method of claim 86, wherein said expression is monitored by assaying nucleic acid level.

25 91. The method of claim 86, wherein said nucleic acid is in a nematode.

92. A method for identifying a candidate compound that treats a neoplasia, said method comprising:

- (a) providing a cell expressing a a Class C synMuv polypeptide selected from the group consisting of TRR-1, HAT-1, EPC-1, and SSL-1 polypeptide;
- 5 (b) contacting said cell with a candidate compound; and
- (c) comparing the expression of said polypeptide in said cell contacted with said candidate compound to a control cell not contacted with said candidate compound, wherein an increase in the expression of said polypeptide identifies said candidate compound as a candidate compound that treats a
- 10 neoplasia.

93. The method of claim 92, wherein said cell is in a nematode.

94. The method of claim 92, wherein said expression is monitored

15 with an immunological assay.

95. A method for identifying a candidate compound that treats a neoplasia, said method comprising:

- (a) providing a cell expressing a Class C synMuv polypeptide selected
- 20 from the group consisting of TRR-1, HAT-1, EPC-1, and SSL-1;
- (b) contacting said cell with a candidate compound; and
- (c) comparing the biological activity of said polypeptide in said cell contacted with said candidate compound to a control cell not contacted with said candidate compound, wherein an increase in the biological activity of said
- 25 polypeptide identifies said candidate compound as a candidate compound that treats a neoplasia.

96. The method of claim 95, wherein said cell is in a nematode.

97. The method of claim 95, wherein said biological activity is monitored with an enzymatic assay.

98. The method of claim 95, wherein said biological activity is monitored with an immunological assay.

99. A method of identifying a nucleic acid target of a synMuv Class C polypeptide, said method comprising:

(a) mutagenizing a *C. elegans* comprising a first mutation in a Class C synMuv gene selected from the group consisting of *trr-1*, *hat-1*, *epc-1*, and *ssl-1* and a second mutation in a Class A or Class B synMuv gene;

(b) allowing said *C. elegans* to reproduce;

(c) selecting a *C. elegans* comprising a mutation that suppresses a synMuv phenotype; wherein said mutation identifies a nucleic acid target of a synMuv class C polypeptide.

100. The method of claim 99, wherein said second mutation is in a class A synMuv gene.

101. The method of claim 31, wherein said second mutation is in a Class B synMuv gene.

102. A method for identifying a nucleic acid target of a synMuv Class C polypeptide, said method comprising:

(a) providing a *C. elegans* comprising a mutations in a Class C synMuv gene selected from the group consisting of *trr-1*, *hat-1*, *epc-1*, and *ssl-1*;

(b) growing said *C. elegans* on bacteria expressing a dsRNA; and

(c) identifying a dsRNA that suppresses a synMuv phenotype; wherein said dsRNA identifies a nucleic acid target of a synMuv class C polypeptide.

103. A method for identifying a nucleic acid target of a synMuv class C polypeptide, said method comprising:

(a) providing a *C. elegans* comprising mutations in a Class C synMuv gene selected from the group consisting of *trr-1*, *hat-1*, *epc-1*, and *ssl-1* and in a Class A or Class B synMuv gene;

(b) growing said *C. elegans* on bacteria expressing a dsRNA; and

(c) identifying a dsRNA that suppresses a synMuv phenotype; wherein said dsRNA identifies a nucleic acid target of a synMuv class C polypeptide.

104. A method of identifying a nucleic acid whose expression is modulated by a synMuv class C polypeptide, said method comprising:

(a) providing a microarray comprising fragments of nematode nucleic acids;

(b) contacting said microarray with detectably labeled nucleic acids derived from a nematode comprising a mutation in a Class C synMuv gene selected from the group consisting of *trr-1*, *hat-1*, *epc-1*, and *ssl-1* gene;

(c) detecting an alteration in the expression of at least one nucleic acid of a *C. elegans* comprising a mutation in said synMuv class C gene relative to the expression of said nucleic acid in a control nematode, wherein an alteration in said expression identifies said nucleic acid as a nucleic acid modulated by a synMuv class C polypeptide.

105. The method of claim 104, wherein said *C. elegans* further comprises a mutation in a synMuv A or synMuv B gene.

106. The method of claim 104, wherein said *C. elegans* further comprises a mutation in a gene that results in a Vulvaless (Vul) phenotype.

107. The method of claim 104, wherein said gene encodes LET-60.

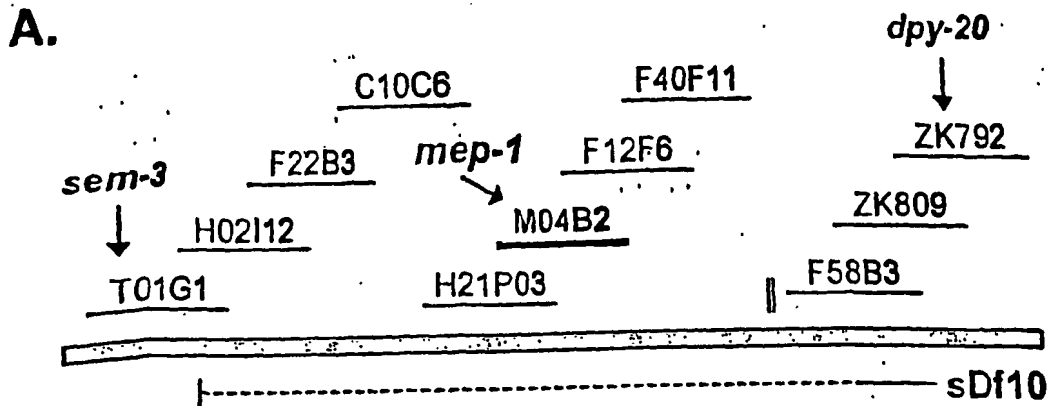
108. A method for identifying a nucleic acid target of a synMuv class C polypeptide, said method comprising:

- (a) providing nucleic acids derived from a nematode cell;
- (b) crosslinking said nucleic acids and their associated proteins to form a
5 nucleic acid-protein complex;
- (c) contacting said nucleic acid-protein complex with an antibody that binds a polypeptide selected from the group consisting of TRR-1, HAT-1, EPC-1, AND SSL-1;
- (d) purifying said nucleic acid-protein complex using an immunological
10 method; and
- (e) isolating said nucleic acid, wherein said isolated nucleic acid is a nucleic acid that binds a synMuv class C polypeptide.

109. The method of claim 108, further comprising the following steps:

- (f) detectably labeling the nucleic acid of step (e);
- (g) contacting said detectably labeled nucleic acid with a microarray
15 comprising *C. elegans* nucleic acid fragments; and
- (h) detecting binding of said detectably labeled nucleic acid, wherein said binding identifies said nucleic acid as a nucleic acid target of a synMuv
20 class C polypeptide.

FIGURE 1

**B.**

```

M V T A D E T V L A T T T N T T S M S V E P T D P R S A G E 30
S S S D S E P D T I E Q L K A E Q R E V M A D A A N G S E V 60
N G N Q E N G K E E A A S A D V E V I E I D D T E E S T D P 90
S P D G S D E N G D A A S T S V P I E E E A R K K D E G A S 120
E V T V A S S E I E Q D D D G D V M E I T E E P N G K S E D 150
T A N G T V T E E V L D E E E P E P S V N G T T E I A T E K 180
E P E D S S M P V E Q N G K G V K R P V E C I E L D D D D D 210
D E I Q E I S T P A P A K K A K I D D V K A T S V P E E D N 240
N E Q A Q K R L L D K L E E Y V K E Q K D Q P S S K S R K V 270
L D T L L G A I N A Q V Q K E P L S V R K L I L D K V L V L 300
P N T I S F P P S Q V C D L L I E H D P E M P L T K V I N R 330
M F G E E R P K L S D S E K R E R A Q L K Q H N P V P N M T 360
K L L V D I G Q D L V Q E A T Y C D I V H A K N L P E V P K 390
N L E T Y K Q V A A Q L K P V W E T L K R K N E P Y K L K M 420
H R L G G V G G G F G L T E E S K K S V M S S G H K K E N L L H 450
Q G E N M G K K E E E D T S E E D R M K D H E Y E E T H L V I A K S E 480
E K E S K Y P C A I C E E D F N F K G V R E Q H Y K Q C K K 510
D Y I R I R N I M M P K Q D D H L Y I N R W L W E R P Q L D 540
P S I L Q Q Q Q Q A A L Q Q A Q Q K K Q Q Q L L H Q Q Q A A 570
Q A A A A A Q L L R K Q Q L Q Q Q Q Q Q Q Q A R L R E Q Q Q 600
A A Q F R Q V A Q L L Q Q Q S A Q A Q R A Q Q N Q G N V N H 630
N T L I A A M Q A S L R R G G Q Q G N S L A V S Q L L Q K Q 660
M A A L K S Q Q G A Q Q L Q A A V N S M R S Q N S Q K T P T 690
H R T P T F V G E E F G C D A S S V G E E K E K Y L L G H E P G F A T H K 720
Q M V G K V L Q D M S G G A P L A K S H G R D R F P V A T O R E G 750
D E E R H L E V M S E H G L V T A D L L L K A Q K K E D G G R I C K 780
E T G C G K N I Y A G F E N M G G H E G V A A D D H I Q V K L C S A E I M Y S 810
G D V A G A F E F G C S S Y G F I L E P A P H U T S E N H P K G D K K T S 840
T P A K K D D C I T L D D 853

```

FIGURE 2

mep-1 genomic sequence:

TCACACACTCATGACATACACACATCATTTGCGCTCACACACCGCGCCGTGCG
CCATCCGCACCGCCCCGGGTGGGACGTGTTCAAACCTTTTCGGTTTTTCGTAAT
TAATAGTGAGCCCCGGTTTATTGCTTTGAGAATCAGTATAATGGATATATG
AGATTGTGTAATTAGGTTGCGTGCTTGAACCTTTAAAATTAACGTGTTTTAAAT
TTATCTGCCTTTATCGTTACAGTAAATCATTTTGATGAACCTTTTCGGATGAAT
CATAATGAAGTACGCAGCGCTCTAACAAAATGTGTTTGTAATTCGAATTGC
TACAAGTTGCCCGGCTTATTTTTTGGTGATTGAAGCATGATTCTGTTGACGC
TCCCGACGCGGAATACCAGGACGGACCGATGAGAGAGTACTGCCAGTGAA
GAGACGCATGCGAGCAGGACGAGTGCTGCTCACCTTCTTCTCAGCGTCG
GCGGCTGCGACCAGCGGCCGAGGAAGGGGAGGAGAGAGGCCGATTGCGC
TGCGTACCACGTTTGATACTCAGTCACTTACCACAGCTGGTTCTCTTGTCG
TTCAAATCTGGCTTGCCGCGCGCGCGCATTATTCCTACCAGTTTGAATCT
CCCACCTCTCCGACTGTAAGTGTCTAATTTGCTTCCTTCTCATCACTCTCTC
TTTGCCTATTTCTCACTATCTAGACTCTATTTTTCCAGATGTCACCGCCGA
CGAGACGGTACTCGCCACAACGACCAACACCACTTCCATGTCTGTGGAACC
AACGGATCCGAGAAGCGCTGGTGAATCGTCCTCAGATTCGGAGCCAGACA
CAATTGAGGTGAGGAAAAGTTTTGGGAATTTAAATCTGAATAAAACGTTTTCA
GCAGCTGAAGGCAGAACAGCGCGAAGTGATGGCCGACGCGGCGAATGGTT
CCGAAGTCAACGGAAATCAAGAGAACGGAAAAGAGGAAGCGGCATCTGCA
GACGTGGAAGTGATCGAGATAGATGACACCGAAGAGTCTACGGATCCCTCA
CCTGATGGATCTGATGAAAACGGTGATGCTGCATCTACATCGGTTCCAATC
GAAGAGGAAGCGCGTAAAAAGGATGAGGGGGCTTCCGAAGTGAAGTGTGGC
ATCATCTGAGATTGAACAAGACGATGATGGCGATGTTATGGAAATCACTGAG
GAGCCGAACGGAAAGTCCGAGGATACTGCCAACGGAACAGGTGTGTTTTAT
AATTTTACCAAGTTTAATTTTAACTTTCTATTTTCAGTTACTGAGGAGGTGCTA
GATGAAGAGGAGCCAGAACCTTCCGTAAACGGAACTGAGATCGCTACA
GAGAAAGAGCCAGAAGATTCTTCAATGCCTGTGGAACAGAATGGGAAGGGT
GTGAAGCGGCCTGTGGAATGCATCGAACTCGACGACGACGATGATGACGA
GATTCAGGAAATTTCTACCCCTGCCCCAGCTAAAAAAGCTAAAATTGATGAT
GTCAAGGCGACAAGCGTTCCAGAAGAGGACAACAATGAGCAGGCGCAGAA
GAGATTGCTCGACAAGCTGGAAGAGTATGTGAAGGAGCAGAAGGATCAACC
ATCCAGCAAAAGCCGAAAAGTTCTGGACACTCTTCTCGGAGCAATCAATGC
GCAAGTTCAAAAGGAGCCTCTGTGCGTTCCGGAAGCTGATCCTGGACAAAGT
TCTCGTTCTCCCAACACAATATCATTCCCACCAAGTCAAGTTTGCGACTTAT
TGATTGAGCACGATCCCGAAATGCCTTTGACGAAGGTTATCAACAGGATGTT
TGGAGAAGAAAGACCAAGTTGAGTGATTCCGAGAAACGAGAGAGAGCTCA
GCTGAAACAACATAATCCTGTTCCAAATATGACAAAACCTGCTCGTGGACATT
GGACAGGATCTCGTTCAAGAAGCTACCTATTGTGATATAGTTACGCGGAAGA
ATCTTCCAGAGGTGCCAAAAAATCTTGAAACCTATAAGCAAGTCGCTGCGCA
GTTGAAACCAGTTTGGGAGACATTGAAACGCAAAATGAGCGGTACAAGTT
GAAAATGCATCGATGCGACGTCTGTGGATTCCAGACGGAATCAAAGCTGGT
TATGAGCACTACAAGGAGAATTTGCACTTCACAGGATCCAAATTCCAGTGC
ACCATGTGTAAAGAGACGGACACGAGTGAGCAAAGAATGAAGGATCACTAC
TTGTAAGTTTTTTTTTTTTCATCTTTCAATATTCATTTAATTACAGCGAAACTC
ATCTTGTTATTGCAAAATCGGAAGAGAAGGAGTCCAAGTATCCATGTGCAAT

FIGURE 2

CTGCGAAGAAGACTTCAATTTCAAAGGTGTCCGTGAGCAGCATTACAAGCA
GTGCAAGAAGGACTACATTTCGCATTGAAACATCATGATGCCGAAGCAAGA
CGATCATCTCTATATCAACAGATGGCTCTGGGAGAGGCCCAATTGGATCC
CAGCATTCTTCAACAGCAGCAACAAGCTGCTCTTCAGCAAGCTCAACAAAAG
AAGCAACAGCAACTTCTGCATCAACAGCAAGCAGCACAAAGCTGCAGCCGCT
GCGCAACTCTTACGGAAGCAACAATTACAACAGCAACAACAACAGCAACAG
GCTCGTCTTCGTGAGCAACAGCAAGCGGCCCAATTCCGGCAAGTGGCTCAA
CTGCTGCAACAACAATCAGCGCAGGCTCAACGTGCACAGCAGAATCAAGGA
AATGTGAATCATAACACTCTGATTGCAGGTAATAGCTAAACATATTTTAAATA
AGTATTTTGTATAATTATTTATATTTTCAGCAATGCAAGCGTCGTTGCGTAGAG
GTGGTCAACAAGGAAATTCGCTGGCAGTTTCTCAACTTCTCCAAAAGCAAAT
GGCAGCTTTGAAGTCGCAACAAGGAGCTCAACAACCTTCAGGCTGCGGTGAA
CTCCATGAGAAGCCAGAACAGTCAAAAGACGCCAACACACAGAAGTTCGAA
ACTTGTTACTACGCCGTCTCATGCTACTGTTGGCTCTTCTTCAGCTCCCACG
TTTGTATGCGAAATTTGTGATGCGTCAGTGCAGGAAAAGGAGAAGTATCTAC
AGCATCTTCAGGTAATTTTAAAGAAACGTTTCTATTTCAATTTCAAACCGATT
ATTAAATATCTTAAACATCACATTTTCAGACTACTCATAAGCAGATGGTTGGA
AAAGTGCTGCAGGACATGTCGCAAGGAGCTCCACTGGCATGTTCTCGATGC
CGTGACAGATTCTGGACTTATGAAGGGTTGGAGCGGCACCTTGGTGATGTG
CATGGTCTCGTCACTGCTGATCTGCTCCTCAAAGCGCAAAAGAAGGAAGAC
GGAGGTCGATGCAAGACATGCGGCAAGAACTATGCGTTCAACATGCTTCAA
CACTTGGTAGCTGATCATCAAGTGAAGTTGTGCTCGGCTGAAATCATGTACT
CGTGCGATGTGTGCGCGTTCAAATGCTCGAGTTATCAGACTCTGGAAGCCC
ATCTCACTTCAAATCACCCAAAAGGAGATAAGAAGACATCAACACCAGCAAA
AAAAGATGATTGTATTACTCTGGATGATTAAATAGGAAAACGAATGGCTTATC
CCGTTCTACGAATGAGTGCTGGAAACATTCTTCACAATGATCTCAATTATTT
TCTTATTCTTTACATTCAATCATTTTAAATCACCAAGTTCTCCCACTTTTATTGA
TATACACATTCTATTGCGGGTTCCGGAACCGAAATCAATCAGTACTTTACTTT
ATTTCCCAATTTTTCTCTTCATGATATCTGGTTTATTCTCGCATCTTCCCCTA
CCTTCAAAACTCCCTATTTTTTTTTTCAAACCTAACTACCCACAAATTATCATG
TAAATCAAATTGCAATCCCCATAAGACAGATCAGTATACACTTTCACTTCA
TACGTCTGTTGTTCTCCCCATCTCATACTTTTTTTACCATTGTCCAGTTAA
GATTTTTGGAAGATATCTAT

FIGURE 3

mep-1 ORF

ATGGTCACCGCCGACGAGACGGTACTCGCCACAACGACCAACACCACTTCC
ATGTCTGTGGAACCAACGGATCCGAGAAGCGCTGGTGAATCGTCCTCAGAT
TCGGAGCCAGACACAATTGAGCAGCTGAAGGCAGAACAGCGCGGAAGTGAT
GGCCGACGCGGCGAATGGTTCCGAAGTCAACGGAAATCAAGAGAACGGAA
AAGAGGAAGCGGCATCTGCAGACGTGGAAGTGATCGAGATAGATGACACC
GAAGAGTCTACGGATCCCTCACCTGATGGATCTGATGAAAACGGTGATGCT
GCATCTACATCGGTTCCAATCGAAGAGGAAGCGCGTAAAAAGGATGAGGGG
GCTTCCGAAGTGAAGTGTGGCATCATCTGAGATTGAACAAGACGATGATGGC
GATGTTATGGAAATCACTGAGGAGCCGAACGGAAAGTCGGAGGATACTGCC
AACGGAACAGTTACTGAGGAGGTGCTAGATGAAGAGGAGCCAGAACCCTTCC
GTAAACGGAACAACCTGAGATCGCTACAGAGAAAGAGCCAGAAGATTCTTCA
ATGCCTGTCGAACAGAATGGGAAGGGTGTGAAGCGGCCTGTCGAATGCAT
CGAACTCGACGACGACGATGATGACGAGATTGAGGAAATTTCTACCCCTGC
CCCAGCTAAAAAAGCTAAAATTGATGATGTCAAGGCGACAAGCGTTCCAGA
AGAGGACAACAATGAGCAGGCGCAGAAGAGATTGCTCGACAAGCTGGAAG
AGTATGTGAAGGAGCAGAAGGATCAACCATCCAGCAAAAGCCGAAAAGTTC
TGGACACTCTTCTCGGAGCAATCAATGCGCAAGTTCAAAAGGAGCCTCTGT
CGGTTCCGAAGCTGATCCTGGACAAAGTTCTCGTTCTCCCAACACAATATC
ATTCCCACCAAGTCAAGTTTGCGACTTATTGATTGAGCACGATCCCGAAATG
CCTTTGACGAAGGTTATCAACAGGATGTTTGGAGAAGAAAGACCAAAAGTTGA
GTGATTCCGAGAAACGAGAGAGAGCTCAGCTGAAACAACATAATCCTGTTT
CAAATATGACAAAACCTGCTCGTGGACATTGGACAGGATCTCGTTCAAGAAG
CTACCTATTGTGATATAGTTCACGCGAAGAATCTTCCAGAGGTGCCAAAAA
TCTTGAAACCTATAAGCAAGTCGCTGCGCAGTTGAAACCAGTTTGGGAGAC
ATTGAAACGCAAAAATGAGCCGTACAAGTTGAAAATGCATCGATGCGACGT
CTGTGGATTCCAGACCGGAATCAAAGCTGGTTATGAGCACTCACAAGGAGAA
TTTGCATTCACAGGATCCAAATTCCAGTGCACCATGTGTAAAGAGACGGAC
ACGAGTGAGCAAAGAATGAAGGATCACTACTTCGAAACTCATCTTGTTATTG
CAAAATCGGAAGAGAAGGAGTCCAAGTATCCATGTGCAATCTGCGAAGAAG
ACTTCAATTTCAAAGGTGTCCGTGAGCAGCATTACAAGCAGTGCAAGAAGG
ACTACATTCGCATTGAAACATCATGATGCCGAAGCAAGACGATCATCTCTA
TATCAACAGATGGCTCTGGGAGAGGGCCCCAATTGGATCCCAGCATTCTTCA
ACAGCAGCAACAAGCTGCTCTTCAGCAAGCTCAACAAAAGAAGCAACAGCA
ACTTCTGCATCAACAGCAAGCAGCACAAGCTGCAGCCGCTGCGCAACTCTT
ACGGAAGCAACAATTACAACAGCAACAACAACAGCAACAGGCTCGTCTTCG
TGAGCAACAGCAAGCGGCCCAATTCCGGCAAGTGGCTCAACTGCTGCAACA
ACAATCAGCGCAGGCTCAACGTGCACAGCAGAATCAAGGAAATGTGAATCA
TAACACTCTGATTGCAGCAATGCAAGCGTCGTTGCGTAGAGGTGGTCAACA
AGGAAATTCGCTGGCAGTTTCTCAACTTCTCCAAAAGCAAATGGCAGCTTTG
AAGTCGCAACAAGGAGCTCAACAACCTTCAGGCTGCGGTGAAGTCCATGAGA
AGCCAGAACAGTCAAAGACGCCAACACACAGAACTCCCACGTTTGTATGC
GAAATTTGTGATGCGTCAGTGCAGGAAAAGGAGAAGTATCTACAGCATCTTC
AGACTACTATAAGCAGATGGTTGGAAAAGTGCTGCAGGACATGTGCGCAAG
GAGCTCCACTGGCATGTTCTCGATGCCGTGACAGATTCTGGACTTATGAAG
GGTTGGAGCGGCACTTGGTGATGTCGCATGGTCTCGTCACTGCTGATCTGC

FIGURE 3

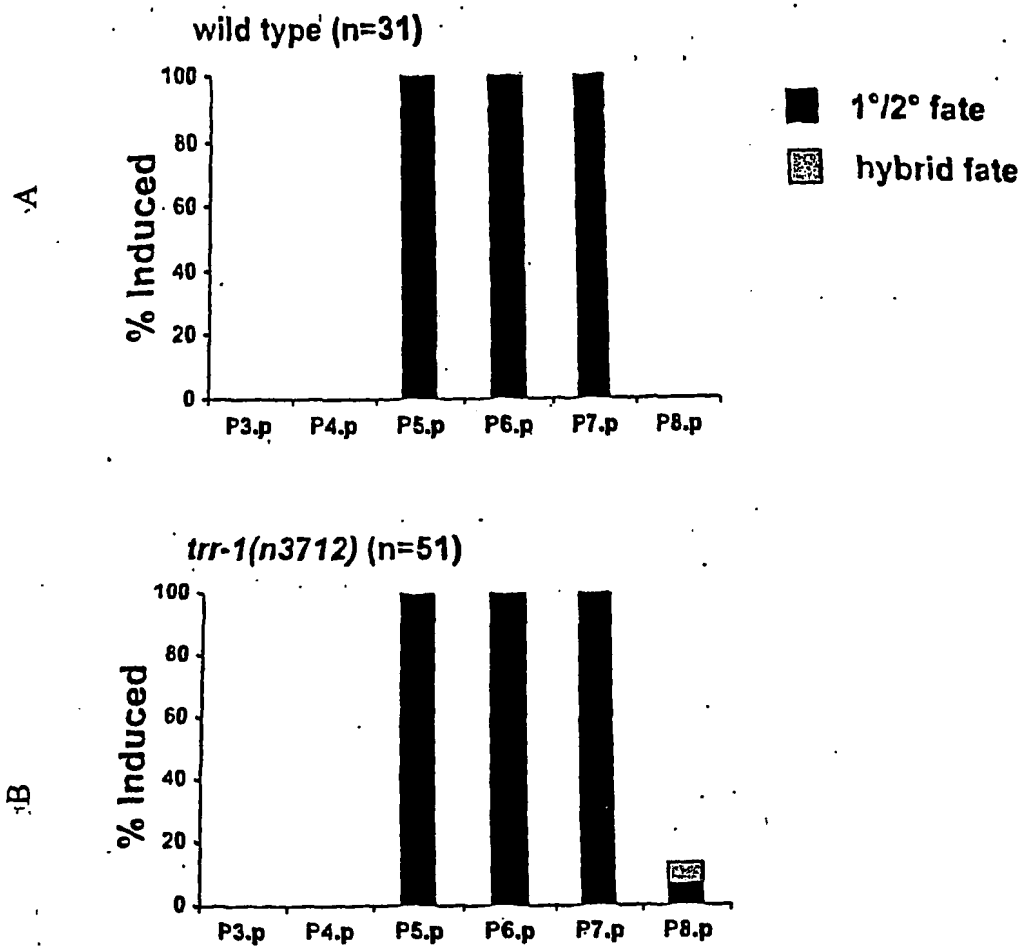
TCCTCAAAGCGCAAAAGAAGGAAGACGGAGGTCGATGCAAGACATGCGGC
AAGAACTATGCGTTCAACATGCTTCAACACTTGGTAGCTGATCATCAAGTGA
AGTTGTGCTCGGCTGAAATCATGTACTCGTGCGATGTGTGCGCGTTCAAAT
GCTCGAGTTATCAGACTCTGGAAGCCCATCTCACTTCAAATCACCCAAAAGG
AGATAAGAAGACATCAACACCAGCAAAAAAAGATGATTGTATTACTCTGGAT
GATTAA

FIGURE 4

MEP-1 protein

MYTADETVLATTTNTTSMŠVEPTDPRSAGESSSDSEPDTIEQLKAEQREVMAD
AANGSEVNGNQENGKEEAASADVEVIEIDDEESTDPSPDGSDENGDAASTSV
PIEEEARKKDEGASEVTVASSEIEQDDDGDMVEITEEPNGKSEDTANGTVTEEV
LDEEEPEPSVNGTTEIATEKEPEDSSMPVEQNGKGVKRPVECIELDDDDDDDEIQ
EISTPAPAKKAKIDDVKATSVPEEDNNEQAQKRLLDKLEEVKEQKDQPSSKSR
KVLDTLLGAINAQVQKEPLSVRKLILDKVLVLPNTISFPPSQVCDLLIEHDPEMPL
TKVINRMFGEERPKLSDSEKRERAQLKQHNPPVPMNTKLLVDIGQDLVQEATYC
DIVHAKNLPEVPKNLETYKQVAAQLKPWWETLKRKNPEPYKLKMHRCDVCGFQT
ESKLVMSHKNLHFTGSKFQCTMCKETDTSEQRMKDHYFETHLVIKSEEKE
SKYPCAICEEDFNFKGVREQHYKQCKKDYIRIRNIMMPKQDDHLYINRWLWER
PQLDPSILQOQQQQAALQQAQKKQOQLLHQQQAAQAAAAAQLLRKQQLQQQ
QOQQQARLREQQQAQFRQVAQLLQOQSAQAQRAQQNQGNVNHNTLIAAM
QASLRGGGOQGNLAVSOLLQKQMAALKSOOGAQQQLQAAVNSMRSQNSQKT
PTHRTPTFVCEICDASVQEKEKYLQHLQTTTHKQMVGVLDMSQGAPLACSR
CRDRFWTYEGLERHLVMSHGLVTADLLLKAQKKEDGGRCKTCGKNYAFNMLQ
HLVADHQVKLCSAEIMYSCDVCAFKCSSYQTLEAHLTSNHPKGDKKTSTPAKK
DDCITLDD

FIGURE 5



8/92

FIGURE 6

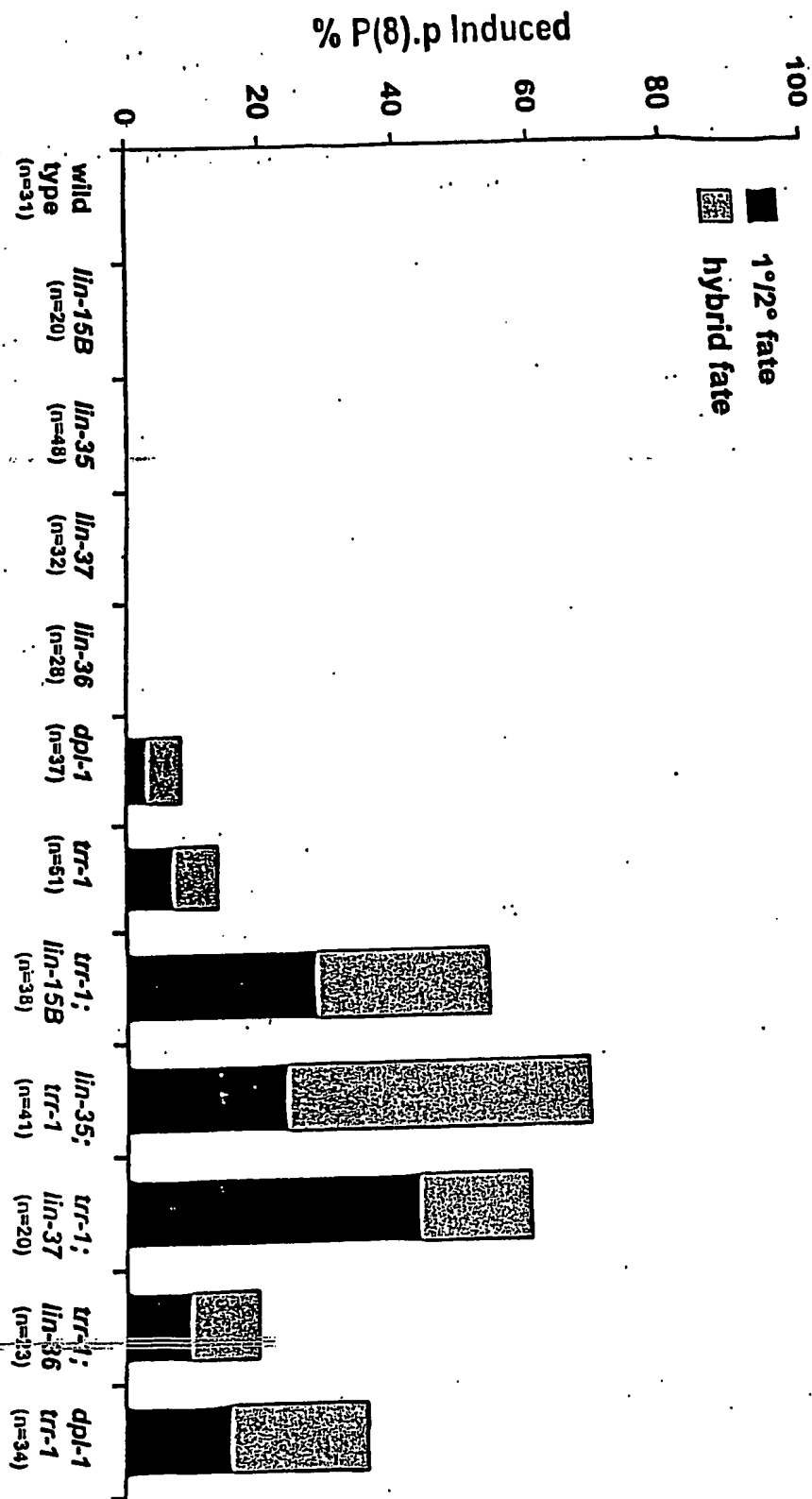
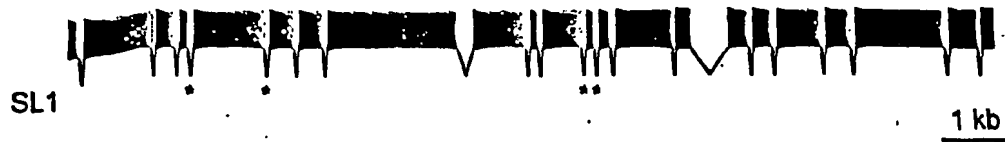
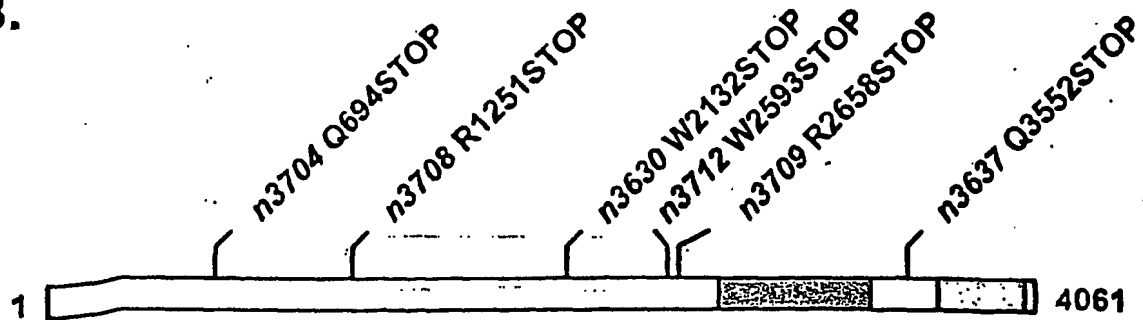


FIGURE 7

A.



B.



■ FAT domain (FRAP, ATM, TRRAP-like)
□ ATM/PI-3 kinase-like

10/92

FIGURE 8

lrr-1 genomic sequence:

GAGGAA GATGTAGACGACGATTCGGTTTCCGTACTCTCATGACTTTTGGCG
AAAATCCTCACGAATTCCTTTTCCGTCATACGTTGAGTTAAAAATCTGGCGAT
GTAACG AAGAATGAGAAGAGCGTTTGATGTTGCCATAAGTAGATTTTACTG
AAATAAGAAAAAGCTTTAATTAATATAATGATGATTTTTTTTCCAACCTCACT
TTTCGCATTGTTCTGATGTTTTAGTTCTGTGGCTCTGCGAAGGAAAAGTCG
AATAAATGCAGCGAAATTCCTGTTGTTTGTGTATTGTACATTAGACATTGAA
GATGATCATCTAAAGCAGATTCCAAAGCGATTCCGGTGTCTCTAAACGATTA
TAACATTTTTAAAGCTTTTGCCTAATTTAATCCTTACTCGTCGTCATCATCAA
ACTTGAGACTGAAAGAGAGAAGTTTGTTCCAAAATGGGTCATAATCGTCGAC
AGGTTCCAAACCGCTGAGTTTCTTCAGATAAATAATTCTCCTGTAAGACCGTT
TCCTTGGTTATAACTGATCCCATGTGTCTGAAATTTGTTATTACACTGTTAAT
AATCATAAAAATAAAAGAAAAAGTCAAGAAAGGGTCAAATATTAATCAGGTCA
CATCTTTTTTATTCAATAAATCTCCTCTCTCGTTCTGGCAATGCACGTGAA
ATGCGCCAACAACCGCGAGTGCGCCAACACACACACATACGCGTCAGCAG
ACAATTCGCTCTCGTTTGAAATTTAGTTGTTTCTTTGTTTCTGCTGAAATAAT
GTCAGTTTTCCGATAATTCAGCGTTTTCTGACTGATTTTTCTTGTTGCATTC
ACTTCCTAATAGTTCATTCTACTCCATTCTTCATTTTATAATCTGTTTCCTTCG
CAATTTAGTGAATTAACACGTAAATCTTGTTTCAGATAAATTATTCAAATAGT
TGCACAAAGCTCAATAGTTTAGAAGTATCTTCAGTGCTGGTCACTAATACAA
AATG GATCCGGCTATGGCTTCTCCAGGCTATCGGTCTGTGCAGTCCGATCG
GAGTAATCACCTAACAGAGCTGGAAACGAGAATTCAAATCTTGCCGATAAT
TCACAAAGAGATGATGTCAAATTGAAAATGTTACAAGTTAGTTTCAATAATTC
GTGTTAAGTAATCAATTTGTTCCGGTTGCAGGAGATTTGGAGCACAAATCGAAA
ATCATTTCACACTAAGTTCGCACGAGAAAGTCGTGGAGAGGGCTCATTCTCTC
GTTCTTACAAGTTTTCTGCAACACAAGTCCACAGTTCATTGCTGAAAACAAT
ACACAACAGCTTCGAAAGTTAATGCTTGAAATCATTCTTCGACTTTCGAACG
TAGAAGCCATGAAACATCATAGCAAAGAAATTATCAAGCAGATGATGAGGCT
AATCACCGTGGAATGAGGAGAATGCCAATTTGGCTATCAAATTTGTCACC
GATCAAGGGAGAAGTACCGGCAAAATGCAATATTGCGGAGAGGTTTCACAG
ATAATGGTCTCCTTCAAACAATGGTCATTGATCTGACGGCGAGTGGTCGA
GCTGGTGATATGTTCAACATAAAAGAGCATAAAGCTCCACCGTCAACTAGCT
CCGACGAGCAAGTCATCACTGAATATTTGAAGACTTGCTACTATCAACAAAC
GGTTCTTCTCAACGGAACGGAAGGAAAACCGCCATTAAAATACAATATGATT
CCATCAGCTCATCAGTCAACGAAGGTGCTCCTGGAGGTTCCGTATCTCGTG
ATTTTCTTCTATCAACATTTCAAACAGCGATCCAAACCGAAGCGCTTGATT
CATGAGGCTTGGTCTTGATTTCTAAATGTCAGAGTTCCAGACGAGGATAAA
CTCAAAACAAATCAAATAATAACCGATGATTTTGTCAGTGCACAGTCCCGAT
TCCTGTCAATTCGTCAACATTATGGCTAAGATTCAGCGGTAAAGTTTCGTTTTT
TCAAGTTTTTTTCTGTAATCCIGATTTTTATTTTTAGTTTTATGGATCTTATCA
TGCAAAATGGACCGCTTCTAGTGTGCGGAACAATGCAGATGCTCGAGCGGT
GCCCCGGCTGATCTGATAAGTGTCCGACGAGAAGTTCTGATGGCTTTGAAGT
ATTTACATCTGGAGAAATGAAGTCGAAATCTTTCCAATGCTACCTCGACT
CATCGCTGAGGAGGTTGTTCTGGGAACAGGATTCACTGCGATTGAGCATTT
GCGAGTTTTCATGTATCAAATGCTAGCAGATCTGTTGCATCACATGCGAAAT
TCTATAGACTATGAAATGATCACACAGTAAGTTTGAATAAGACTTTCTGATGA

FIGURE 8

AAAATGTTGAAATTT CAGCGTGATTTTCGTATTCTGTCGCACTCTTCACGATC
CTAACAACCTCTTCTCAAGTCCAGATTATGTCTGCTCGGCTGCTCAACTCACT
GGCCGAATCTCTGTGCAAAATGATTCAATGATACCGTAAGACTTATTCTA
TCAATAATCGTATCTCACTTCGAAATAAGTTTCAGACTCGTGATCTGCTCATT
GAAATCCTGGAGTCGCACGTGGCCAAGCTCAAACTCTTGCAGTCTATCAC
ATGCCATTCTCTTCCAACAATACGGAACCGAAATAGACTACGAATACAAAA
GTTATGAGAGAGACGCCGAGAAACCTGGAATGAATATCCCAAAGGACACTA
TACGAGGAGTACCGAAACGAAGAATCCGTGCGGCTCTCCATTGATTGAGTTG
AAGAGCTGGAATTCCTGGCATCAGAACCATCCACGTGCGAAGATGCAGATG
AGAGTGGTGGAGATCCGAACAAGCTTCCTCCGCCAACAAAAGAGGGAAAGA
AAACGTCTCCCGAAGCGATTTTAACCGCCATGTCAACGATGACACCTCCTC
CATTGGCAATTGTTGAAGCTCGAAATCTTGTGAAGTATATAATGCATACGTG
TAAATTCTGACAGGACAATTGAGAATCGCCCGGCCATCACAGGATATGTAT
CATTGTTTGAAGGAGCGAGATTTATTGCAACGTCTTCTACGATATGGTGTAA
TGTGTATGGATGTATTCTGTGCTTCCAACAACCTCGAAATCAACCACAAATGCA
TTCTTCAATGCGGACAAAAGATGAGAAAGATGCTCTGGAGTCGTTGGCAAA
CGTTTTTACAACAATCGACCATGCGATATTCCGGGAAATCTTCAAAAAGTAT
ATGGATTTCTTGATTGAAAGAATTTACAATCGGAACCTATCCATTGCAATTGAT
GGTGAACACCTTCTTGTTTGAAGTGAAGTGCCATTCTTCGCATCTACGATG
CTTTCATTCTTGATGTCTCGAATGAAATTGCTGGAAGTTAGCAATGACAAGA
CGATGCTATATGTGAAGCTCTTCAAAATTATCTTCTCCGCCATCGGAGCCAA
TGGCTCTGGGCTTCATGGAGATAAAATGCTCACTTCATACCTCCAGAGATT
CTCAAACAGTCAACTGTCTTGGCATTAAACAGCTCGTGAACCTCTCAACTATT
TCCTTTTGCTTCGTGCAATTGTTCCGCAGTATTGGTGGTGGCGCTCAGGATAT
TTTGTATGGAAAGTTCTGCAGTTACTGCCAAATCTTCTTCAATTCTTGAATA
AATTGACGGTGAGTTTCATTTTTTATATATCGGTAATACACTAAAAATCCAG
AATCTTCAGTCATGTCAACATCGGATTCAAATGCGTGAGCTCTTCGTGCGAGT
TGTGTTTGACTGTGCCAGTTCGACTCAGTTCCCTTCTGCCATACCTACCGCT
TCTGATGGATCCACTGGTGTGTGCGATGAATGGGAGTCCGAACATAGTTAC
ACAAGGATTGAGAACATTGGAATTATGTGTGGATAACTTGCAACCTGAATAT
CTTCTCGAAAATATGCTTCCTGTCCGTGGAGCTTTGATGCAAGGCCTCTGG
CGTGTGTATCGAAAGCTCCAGATACATCATCGATGACAGCAGCGTTCAGG
ATCCTCGGAAAGTTCCGGAGGAGCCAATCGAAAACCTTCTGAATCAACCGCAA
ATTCTTCAAGTAGCCACTTTAGGCGACGTAAGTTTATTTAGTTTATTCTCTTC
CTCGTTTTAAGTTCTAACATTGATCCTATTAACAGACTGTTGAGTCGTACATC
AATATGGAATTCTCGCGGATGGGACTCGATGGCAATCACAGCATTACCTG
CCACTGTCCGAGTTGATGAGAGTCGTTGCCGATCAGATGAGATATCCAGCT
GATATGATCCTTAATCCAAGTCTTGAATGATCCCGTCAACTCATATGAAGA
AATGGTGTATGGAATTGTGAAAGCCGTCTTGTTAGCCGGACTTGGATCTTC
AGGAAGCCCAATTACTCCAAGTGCAAACTTCCGAAGATTATCAAGAACTT
CTTGAAGATTTTGTATCCAAACAATCGTACCCTGAAGTATACACATGTCCGA
GGGAAAGTGATCGAGAGCTTTTTGTGAATGCACTTCTCGCAATGGCTTGTA
GTTCTTAAGTTCTTTTCTCTAATCAGATCTATATTTTAAATTTTTCAGACGG
AATATGGAATAAAGACGGTTTCCGGCATGTCTATAGCAAATCTTTATCAA
GTTCTCCGCCAGTTTGCGTTGATTGGAGTACTCGAATACATTGGTGGAAATG
GATGGATGCGTCATGCAGAAGAGGAAGGTGTTCTACCATTTGTCCTTGACT

12/92
FIGURE 8

CGTCTGTTATGGTTGATGCTCTGATTATTTGTCTCTCTGAAACATCGTCAAG
CTTCATCATTGCTGGTGTCTCTCTCGTCATATCAATGAGACTCTCTCG
CTTACACTTCCCGATATTGATCAAATGTCGAAAGTTCCAATGTGCAAATAGTT
GATGGAGAAGGTGTTCAAATTGTGTACGGGCCTGCTTGGTATGCAAGATC
TGGTGGAAATCAATGCAATTGGATACATGATCGAATCGTTTCCACGAAAATTT
GTTATGGACTTTGTGATAGATGTTGTTGATTGATCATGGAAGTTATTTTGG
GAACTGTTGAAGAAATATCAAGTGGATCTGCTGATTCTGCATACGATTGTCT
CAAGAAAATGATGCGAGTCTATTTTCATCAAAGAAGAAGGCCAAGAAGAGGA
GAATCTGACACTCGCGACTATTTTTGTGTCTGCAATCTCTAAGCATTACTTCC
ACAGTAATGAAAGAGTCAGAGAATTTGCGATTGGTTTAATGGATCATTGTAT
GGTTCACTCAAGACTTGACCATCCCTTGATAAGTTCTACTATCGATTCAAG
GAGTTCTTTGAGCCAGAATTAATGCGGGTGCTCACAACAGTTCCAACAATGT
CATTGGCAGACGCAGGAGGAAGTTTGGATGGAGTTCAAACTATATGTTCA
ACTGTCCGGATGGTTTTGATTTGAAAAAGATATGGACATGTACAAGCGATA
TTTGTACATCTGCTGGATATTGCACAAACCGATACATTTACCTTAAACCAA
GGAATGCCCTTCAAAAAATGCGAGACATGCCCATCGCATTTCCTTCTCCATT
CCCAATCACTACACATATTGATTCAATGCGAGCCAGTGCTCTACAGTGTCTT
GTGATCGCGTATGATCGAATGAAGAAGCAATACATCGACAAGGGAATAGAG
CTGGGTGATGAGCATAAGATGATAGAGATCCTCGCACTTCGCAGCTCCAAG
ATCACAGTTGATCAAGTCTACGAGAGCGATGAATCTTGGAGACGATTGATGA
CAGTTCTATTGAGAGCAGTCACTGACAGAGAACTCCTGAAATTGCGGAGA
AGCTTCATCCTTCACTTTTGAAGGTCTCACC AATATCCACAATCATCATCGCA
ACATTTGGTGCTTCTTACATAAGAAATATTAGTGGAGCAGGAGATGACAGTG
ATTCAGATCGTCATATTTCGTACAACGATATAATGAAGTTCAAGTGTCTCGTG
GAGCTCAATCCAAAGATTCTGGTCACAAAAATGGCAGTGAATCTCGCAAATC
AAATGGTTAAATATAAGATGAGTGACAAGATCTCTAGGATTTTGTGAGTTCC
CAGTAGCTTCACTGAAGAGGAGCTCGATGATTTCGAAGCGGAGAAGATGAA
AGGAATTCGAGAGTTGGATATGATTGGTCATACGGTTAAAATGCTTGCTGGA
TGCCCAAGTGACCACATTCACGGAGCAAATTATTGTGGATATCAGTCGTTTTG
CTGCTCATTITGAGTATGCTTATTCGCAAGATGTACTTGTAATTGGATTGAT
GATGTCACAGTAATCCTCAACAAAAGTCCCAAAGATGTATGGAAGTTCTTCT
TGTCTCGAGAATCAATTCTAGATCCTGCACGCAGATCCTTTATTGGAAGAAT
CATAGTCTATCAATCAAGTGGTCCACTGCGACAGGAATTCATGGATACTCCG
GAATATTTTGAAGAACTCATTGATCTTGACGATGAGGAGAATAAGGATGAAG
ATGAGAGAAAAATCTGGGATCGTGATATGTTTGCATTTTCGATTGTGATCG
TATCTCGAAGAGCTGCCCTGAGTGGCTTATTTCTCCGAATTCCCCAATTCCA
AGAATTAAGAAGTTGTTCTCCGAAACGGAATTC AATGAGCGATATGTGGTTC
GAGCATTGACTGAGGTGAAGAAATTTCAAGAAGAGATCATAGTGAAACGGA
TGACAGAGCACAAGTACAAGGTTCCGAAGCTGATTCTGAATACCTTCTGA
GATATTTGAGGTAATTTCAAGATAGTTTGTAAAAATTAATTACAAAGAAATATA
CCAAAATGAACCCCAAAAAAATTTTTGAATTTGCGATCAAAAAAATTTAA
TATTTTCTCGAAAAATCCTTCAAAATACCAAAAAATTCGAATTCCTCACTTCTAA
AATTATTTTGAATTTTAAATAATTTTTGAACATTTCTCTATGAAATTCATGTT
TTGGGCCTATTTCAAGGCTATAAAAAATTTTTTCTGATTTTAAATAACTTGCAA
ATTCAGGCTCAACATCTATGACTACGATCTATTCATCGTTATCGCCTCGTGT
TTCAATGGCAATTTGTCACCGATCTCTCTTTTCTTCGCGAATATCTTGAAAC

13/92
FIGURE 8

TGAAGTCATCCCGAAAGTGCCGTTACAATGGCGGAGAGAGCTGTTTCTTCG
AATTATGCAGAAGTTTGATACGGATCCACAACTGCTGGAACAAGTATGCAG
CATGTGAAGGCCCTTCAATATTTGGTTATTCCCACGTTGCATTGGGCGTTCCG
AGCGATATGATACGGATGAAATTGTTGGCACCGCACCAATAGATGATTCCG
ATTCTTCGATGGATGTAGATCCGGCAGGCAGCTCGGATAACCTTGTGGCTC
GTTTAACATCAGTCATTGATTCTCATCGTAATTATCTGAGCGATGGAATGGT
CATTGTTTTCTATCAACTTTGCACATTGTTCTGTACAAAACGCCTCCGAACATA
TTCACAATAATAACTGCAAGAAACAAGGTGGACGCCTACGGATCCTGATGCT
CTTCGCCCTGGCCGTGCCTGACCATGTACAATCATCAAGATCCAACAATGCG
GTACACTGGATTCTTCTTCTTGGCCAATATTATAGAGCGTTTCACAATTAATC
GGAAAATCGTGCTTCAAGTGTTCCATCAACTTATGACTACTTATCAGCAGGA
CACTAGAGATCAAATCCGGAAAGCCATTGATATATTAAGTCCAGCTTTGAGG
ACACGAATGGAAGATGGACACTTGCAAATATTGAGTCATGTGAAGAAAATTC
TTATCGAAGAATGCCATAATTTGCAACATGTTTCAGCATGTTTTGTAAGTTTAT
TATCTAAAATGATTTTTTTTAAATGTTAAAAATTTAATTTTAAAATGCGTTCGTG
CTCCTTTAATAATTCTGAATTTCCAGCCAAATGGTGGTTCGCAATTATCGT
GTCTACTATCATGTTTCGATTGGAGCTTCTCACGCCTCTTCTGAACGGAGTTC
AACGAGCACTTGTGATGCCAAATAGTGTTCTGGAAAAGTAAGTTTCCAGCCC
GTTGTTTCGTAAACTCACCCCTTGTAATATTTAGCTGGCAAACCTCGACGTCA
TGCGGTGGAGATCTGCGAGATGGTCATCAAGTGGGAATTGTTTCAGAACGCT
GAAAACAGATCATATTATCAGTGACGAAGAAGCTCTCGAAGTTGACAAGCAA
TTGGATAAGCTGCGAACAGCTTCATCCACAGATCGTTTCGATTTTCGAGGAG
GCTCATAACAAGAGAGACATGCCTGATGCTCAACGCACGATTATCAAAGAG
CACGCCGATGTGATTGTCAATATGCTTGTCCGATTCTGTATGACGTTCCATC
AGAATTGGGTTCTTCGTCCACTTCTCAAAGTGGGAACCATGGTGTGAGTT
GACCAAAAAATGTCAGCTGCTTCTACGTGCAGCCCTACGACCAAGCATGTG
GGGAGAATTTGTCAGCTTCCGATTAACAATGATCGAAAAGTTTTTGTCAATT
CCGAATGATAATGCTCTACGCAATGATATAAGTTCTACGGCCTACGCTAATA
CTATCCAAAATGCACAACACACTCTGGATATGCTGTGTAATATTATTCCTGTT
ATGCCAAAACTAGCTTGATGACTATGATGAGACAACTCCAACGGCCACTCA
TACAATGTCTCAATAACGGAGCTCAGGTATGTGAAGAACGATGAATAGGGG
GTTATAAATCACTAATTTCTCTTAGAACTTTAAGATGACTCGTCTTGTCACTC
AAATTGTCAGTCGGTTACTCGAAAAGACAAATGTTTCGGTTAACGGGGCTTGA
TGAGCTGGAGCAATTGAATCAATACATTTCCCGATTCTACATGAACATTTT
GGATCTCTTTTGAAGTAAGTTTTATTTTGAATTTCCATCTTTCAACCCCTTCGC
CAGTTGCAGAACTTGAGTGGACCAGTGTTGGGAGTTCTCGGAGCATTTTC
TCTTTTGGCAACAATTTGTGGACACGAGCCAGCATACTTGGATCATTTGATG
CCTTCATTTGTAAAAGTGATGGAGAGAGCTGCAAAAGAGCACTTGGCGTAT
GTTGCCAACTCGCAAGATGGAAATATGGTGAAGAGTAAGTTCTATAAAAAGA
TTCAGATTTTCTAATCCCTTAGATTCTTTCCAGATGTTGCTGAATTGTTGT
GTGCATGCATGGAGCTGGTACGTCCCAGAGTCGATCATATCAGTATGGAGA
TTAAGAGATCAATTGTTGGTGGTATTATCGCGGAGCTGATTATCAAATCGAA
TCACGATAAGATCATCCAGACGTCAGTGAAGCTTCTCGGAGCAATGATTAG
CACGCAGGATATGGAATTTACAATTCTCACTGTTCTTCCGCTACTTGTTCTG
ATCCAATCAATTATTGTGACCAAGTTCAAGAATTGCAAGGATCTGATAGCAG
ACTATCTTGTGTGGTTATTACCGTTTTTGGAGAACAGCGAATATCGGAACTC

14/92

FIGURE 8

GGAAGCTGGATCTCGTCTCTGGGAAGGATTCTTCTGGGGACTCAAGAGTAG
CGATCCTCAAACCCGGGAGAAATTCTCGATAGTTTGGGAGAAGACTTGGCC
ACACATGGCAACAGTAGATATTGCTCATCGAATGAAATATATCATGCAAAAT
CAAGATTGGTCCAAGTTCAAACACGCGTTTTGGTTGAAATTCGCACTTTGGG
GAATGCTACGAACGATTGCCAAACGGCCAACCTGATCCGAATAATAAGAGAA
AGAAAGTGATACTGTTGAACTGTGCAACTCCATGGAGAACAATTGAATATGC
AGCGAAATTGAAGGATCAGCCAATGGAAGTGGAACTGAAATGAAACGAGA
AGAGCCAGAACCGATGGAAGTTGACGAAAAAGACTCGCAAGATGATTCTAA
GGATGCCGGAGAGCCCAAGGAGAAGGAAAAGCTCACATTGGAATTATTGCT
TGCTGGACAACAAGAACTTTTGGATGAAGCTTCCAATTATGATTTTGCGGAT
GCTCTAGATACAGTATCCCAGATTACATTTGCACTTAATGGTAAATTGTTCAA
AGTTTATGAATATTTTTCTTAAAAATCACATTTTCAGAGAATCAAGTGACAA
GCAAGATGTGGGTAGTGTGTTCAAATCATTCTGGAGTTCCTTATCACAAATC
CGAAATCGAAGATTTACGGCGCTAGTCGTTCCGTTTATGAGCAGTGGAGT
GCATAATAATTATCAGACGGGTGTACAGGATAGTGTGCTTGCTGTTTGGCTT
GAAGCTGTTGGTGACGCTGTTCAATTTGCCGTCCAGATTGATTGAGGTACGTT
CTGAAAATGAATGCTGGAAAAAATTCGATTTTCTGTTTAAAAAAGTTAAAA
TTTCCGATTTTTTGAATAGCAAAAAAAGAAAAACATTTATTTTGA AAAAAGA
GTCCTCACCGGAATTTTTAATAATAAATTTAAAAAAGAAAAAAACTAAA
AACTTCAATTTTTGAAAATCAAAAAAATTAACAGAAACAGACGAGGTAAAA
AATTTTAAAAAAGTTCTGTAAAAAATGGAGAATCACAGTTTTCGTTGTCTT
TTCTGAAAAAATTTGAAAAATTA AAAATTAACGATTTTTTGGTTTTAATTTA
AAAAATATACGAAAAAAGACTGAAGA ACTTTTTTGTCAAAAA ACTTGATT
TTGATGAGGGAAAAAGTTCAAAA ACTTGGAGAAATCATCGGAAATTTTAGAA
GATTCAATAAAAATTTCCAAAAA AATTTGAACATTTATGATTTTTGGGTAT
TTTGAAAAATTGAAAAATTACGCTTAATTTTTAGATTAAAAAATCAAAAAA
ACCAACACTCCTTTTGAACTTGACACTTTTGAAACGTTTTTTTTTTTGCAT
AATAAATTTCTCATTTTCAGTTTATCTCATCAAAACACGAATGCTGGCATAACG
GAATCAGGCTTCTCGAGAATCATATATGGACAATTCCAAAGCAACTCAACAA
CACGTTACTCCGAGAAATGAAAGTGGCACCAGGTCTCGCTGGAGATATTGA
GACACTCGAATCTCTTGGAACACTCTACAATGAGATATCAGAGTTTGATCAG
TTCGCTGCAATCTGGGAACGCCGTGCTGTATTTCTTGATACGATGAGAGCA
ATGTCAGCTATGCAATTGGGAGATATGGAATTAGCTCAATCTTATCTGGAAA
AATCAATGAGCAGTACGTATGAACTCTTGCTCCGACAATCAATCGTAAGTT
TGGATCAATCGGTTGTACTTCTCACACAAAATAGTATTCCTTTCAGCAAACAA
CACTTCAAATTCGGAGAAGCATGTTTCTCCGATTATTGACAAAGAATACGAT
CATTGGATGGAGATGTACATCACAAATTGCTCGGAGCTTCTTCAGTGGCAAA
ATGTGGCCGACGTATGCAATGGCAAAGACATGCAACATGTTCGTGGCCTGA
TCAACGCAGCATCTCACATTCCGGACTGGAATGTGGTTCGAGGAGTGTA AAA
GTCAGATAGCTGGATGTATTCCACCAAGTTTCCATTTAGATTACACTCTTTTC
AATTTGATGAGTACTGTTATGGTTAGTTTAAAGTCAAAAAGTGATATATAATTA
TTGTTTAAATTTTTCAGCGAATGAATGAAA ACTCAAGCCCGACACATATGAAG
GAACGATGCAAAATTGCAATTCAAGAGTGCACAGAAGCTCATATTAGTCGTT
GGAGAGCACTTCCGTCA GTTGTTCATATGGTCATGTCAAGATTCTTCAGGC
AATGA ACTTGGTTTCGAGAAATTGAAGAGTCTACAGATATTTCGCATTGCTCTG
CTCGAGGCCCATCAAACAAAGTGGATCAGGCGTTGATGGGCGATATGAAG

15/92

FIGURE 8

GGAAGCTGGATCTCGTCTCTGGGAAGGATTCTTCTGGGGACTCAAGAGTAG
CGATCCTCAAACCCGGGAGAAATTCTCGATAGTTTGGGAGAAGACTTGGCC
ACACATGGCAACAGTAGATATTGCTCATCGAATGAAATATATCATGCAAAAT
CAAGATTGGTCCAAGTTCAAACACGCGTTTTGGTTGAAATTGCGACTTTGGG
GAATGCTACGAACGATTGCCAAACGGCCAACTGATCCGAATAATAAGAGAA
AGAAAGTGATACTGTTGAACTGTGCAACTCCATGGAGAACAATTGAATATGC
AGCGAAATTGAAGGATCAGCCAATGGAAGTGGAACTGAAATGAAACGAGA
AGAGCCAGAACCGATGGAAGTTGACGAAAAAGACTCGCAAGATGATTCTAA
GGATGCCGGAGAGCCCAAGGAGAAGGAAAAGCTCACATTGGAATTATTGCT
TGCTGGACAACAAGAACTTTTGGATGAAGCTTCCAATTATGATTTTGGCGAT
GCTCTAGATACAGTATCCCAGATTACATTTGCACTTAATGGTAAATTGTTCAA
AGTTTATGAATATTTTTCTTAAAAATCACAATTTTCAGAGAATCAAGTGACAA
GCAAGATGTGGGTAGTGTTGTTCAAATCATTCTGGAGTTCCTTATCACAATC
CGAAATCGAAGATTTACGGCGCTAGTCGTTCCGTTTATGAGCAGTGGAGT
GCATAATAATTATCAGACGGGTGTACAGGATAGTGTGCTTGCTGTTTGGCTT
GAAGCTGTTGGTGACGCTGTTTCAATTTGCCGTCCAGATTGATTGAGGTACGTT
CTGAAAATGAATGCTGGAAAAAATTCGATTTTTCTGTTTAAAAAAAGTTAAAA
TTTCCGATTTTTTGAATAGCAAAAAAAAAAGAAAACATTTATTTTGA AAAAAGA
GTCCTCACCGGAATTTTTTAATAAATAAATTTAAAAAAAGAAAAAAAAGTAAA
AACTTCAATTTTTGAAAATCAAAAAAAAAAATTACAGAAACAGACGAGGTAAAA
AATTTTAAAAAAGTTCTGTAAAAAAATGGAGAATCACAGTTTTCGTTGTCTT
TTCTGAAAAAATTTGAAAAATTA AAAATTAACGATTTTTTGGTTTTAATTTA
AAAAAATATACGAAAAAAGACTGAAGAACTTTTTTGTCAAAAAAAGTTGATT
TTGATGAGGGAAAAAGTTCAAAAACTTGGAGAAATCATCGGAAATTTTAGAA
GATTCAATAAAAATTTCCAAAAAATAAATTGAACATTTATGATTTTTGGGTAT
TTTGAAAAATTGAAAAATTACGCTTAATTTTTAGATTAAAAAATCAAAAAAA
ACCAACACTCCTTTTGAAACTTGACACTTTTGAAACGTTTTTTTTTTTGTCAAT
AATAAATTTCTCATTTTCAGTTTATCTCATCAAAACACGAATGCTGGCATACCG
GAATCAGGCTTCTCGAGAATCATATATGGACAATTC AAAGCAACTCAACAA
CACGTTACTCCGAGAAATGAAAGTGGCACCAGGTCTCGCTGGAGATATTGA
GACACTCGAATCTCTTGGAACACTCTACAATGAGATATCAGAGTTTGATCAG
TTGCTGCAATCTGGGAACGCCGTGCTGTATTTCTGATACGATGAGAGCA
ATGTCAGCTATGCAATTGGGAGATATGGAATTAGCTCAATCTTATCTGGAAA
AATCAATGAGCAGTACGTATGAAACTCTTGCTCCGACAATCAATCGTAAGTT
TGGATCAATCGGTTGTACTTCTCACACAAAATAGTATTCCTTTAGCAAAACAA
CACTTCAAATTCGGAGAAGCATGTTTCTCCGATTATTGACAAAGAATACGAT
CATTGGATGGAGATGTACATCACAAATTGCTCGGAGCTTCTTCAGTGGCAAA
ATGTGGCCGACGTATGCAATGGCAAAGACATGCAACATGTTGTTGGCCTGA
TCAACGCAGCATCTCACATTCCGGACTGGAATGTGGTGGAGGAGTGTA AAA
GTCAGATAGCTGGATGTATTCCACCAAGTTTCCATTTAGATTACACTCTTTTC
AATTTGATGAGTACTGTTATGGTTAGTTTAAAGTCAAAAAGTGATATATAATTA
TTGTTTAATTTTTAGCGGAATGAATGAAAAGTCAAGCCCGACACATATGAAG
GAACGATGCAAAATTGCAATTCAGAGTGCACAGAAGCTCATATTAGTCGTT
GGAGAGCACTTCCGTCAGTTGTTTCATATGGTCATGTCAAGATTCTTCAGGC
AATGAACTTGGTTGAGAGAAATTGAAGAGTCTACAGATATTCGCATTGCTCTG
CTCGAGGCCCATCAAACAAAGTGGATCAGGCGTTGATGGGCGATATGAAG

FIGURE 8

TCGTTGATGAAAGTATTCCGAAATAGAACACCAACCACTTCGGATGATATGG
GATTCGTTTCGACTTGGTATGATTGGAGGAATCAGATTCATGGAATGATGCT
TCAAAGATTCGAATATTGGGATAAAGTAGGACTCAACGTCGCTGCAACTGGA
AACCAGTCAATTGTTCCGATTCAATCAATGGCTCAAGCACAGTTGGCCGTAG
CCAAACATGCCAAGAATCTTGGATTCCATAATTTAACGAAAGATCTACTCAA
CAAATTAGCTGGATTGACAGCCATACCGATGATGGATGCTCAAGATAAAGTT
TGCACTTACGGCAAGACACTTCGCGATATGGCAAACAGTGCGGCTGACGAA
AGAGTGAAAAATGAGCTATTGTGTGAAGCGCTTGAAGTTTTGGAAGATGTGC
GAATTGATGATCTACAGAAGGATCAGGTTGCTGCATTGCTTTATCATCGTGC
TAATATTCATTCAAGTTCTTGATCAGTAAGTTTTCAATGCCGAAAAAAATTA
AGTTTTACAAAAATAAATTTAGAGCTGAAAATGCTGACTACACCTTCTCCGC
AGCCTCTCAACTTGTGCACTTGCAAATAGTGTGACAACCACTGGAATCAAG
CTCATGAAAAATTGGGGCCACCATCTTTACAAGAGATTCTTCTCTACGACAG
TTTGCAAGGAAACCGGAAACAACCTTCGGACGGCAGGCTCTCGCTTGTTACT
TCATTGCGGCTCGTGTGGATAACGATATCAAGGCGAGAAAACCGATTGCCA
AGATTTTGTGGCTCTCGAAGCACTTGAATGCGTGTGGATCACATGAAGTGAT
GAATCGGGTTATTAAGAAGCAACTTCATTCACTTAATCTCTTCAATTGGCTTT
ACTGGCTTCCACAATTGGTTACTGATGTTGATATAAACCAAATTCGAATTT
GTTCTGATTCTCTGCAAGGTAAGTTTTGAAATATTTAAATATTTTCAGAATTT
AAATGAAATTCATTTGCAGATGGCTGCTGCTCATCCACTTCAAGTATTTTACC
ACATTCGGGAGGCGAGTTAGCGTTGACGATATTGACTCGGTTCTCGAAGAAG
ATTACACTGATGAGCAAATGTCGATGGATGTTTCGGATGAGGATTGTTTTGC
AGACGATCCACCATTGATAGAATTCTGAAAATATGTCTGAAATATCGTCCAA
CTGATATTCGAGTCTTCCATCGTGTCTCAAAGAACTTGACGAGATGAATGA
GACATGGGTGAACGTCACCTGCGTCATGCGATCTGCCTCAAGGATCAGAT
GTTCAAAGATTTCTCGGAACAAATGGACGCGACGTTCAATGAGATGCAATAT
TCGGAGGATGTGACTATGATGACGTTGAGATGGAGGAAACAGCTGGAAGAA
GACTTGGTGTATTTCCAACAGAATTATAATCTTGATTTCTTGGAGATTCTGTA
CAAGCGAAAGATGATCGTGACGAAGGGATGTATGGGAGTCGAGAAAAGTCA
GATAATGTTGAAAAAGAGCTGAGTCAAGTGTTACAGAGCCGGCCGGCAT
GCAAGATGAATTTGATTTGTCACAAATATGACTAATATGATGGTCTCACAGT
TGGATATTCATGCAGTCGATGCTCCACGCCCTCAGGGATATATTCGTATTGT
TCTCGACTGGATTGAGCGATTGCTCGTCGTTTCGATCGACTTCCACGAAG
AATCCCTCTGGAATCGTCAAGCCCATATCTCGCCAGATTGAGCCATCGTACA
GGATGCATCGAAATGCCATACGATTTGCTCAACGTTTTGCGCGCCAAGAAT
CATACTCTGATGGCTTCCAATCAAACGGGGCAATACATATCCATGCTCTCTC
GATTTGAGCCAACTTTGAGATTGTGATCAAAGGTGGTCAAGTGATAAGAAA
GATCTATATTCGAGGACAAACCGGAAAGAGTGCGGGCGTTTTATCTGAAGAA
ATCTGTGCAGGATGAGCCAACTAACCAGATTCCACAAATGTTCAAACATCTT
GATCACGTTCTACAACCGATAGAGAGTCGGCGGAGAAGACATCTTCATGCT
CCAACAGTGCTGCAGATGAGAGTCGGACAGAAGACGACACTCTACGAAGTT
GCATCCGTTCAACCATATGCAATGCCACCGGATTGTACCAGAACTATCCAG
CATCACAATCGACATTGTTTCATCCATATGATGTGCTGACTGCCACTTTCAAT
GGAAGTTATTATCCGGATGATATGGTATTGCACTTCTTTGAGAGATTCGCCC
AAAGTTCTTCATCCATCGGACAACCTCTTCCAACCTCCGACGAACCAAGATGG
AACAGTTGCTCCGCCACGACTAACGGAAGCTCACCACATCAAGAATATTATT

FIGURE 8

TATGAGTACGTTTGAGAAGCTAGTGTCTAAAATAATAATTAATGTAAAAAAT
TTTCAGAGACTTTGCCCGAGATATGATCCCATTCCGACTTCTCTACGACTAC
CTCACTGCACGATATCCTGATCCGGTTATGTACTATGCAATGAAGAAGCAAT
TGCTGCACAGTCTCGCCGTCCTATCCACAATCGAATATCATTGCAATCTGAC
ACCAATGGGACCTGATCAAATGATGATGACAATGAATACTGGAGTCCTTAGC
AATCCTTCATATAGATTGAAATCCGAGGAGGACGATCACTTCATGATATTC
AACACTTTGGACATGAAGTTCATTCCGATTGACTCCAAATCTATCGATTTTG
GTTGGTGTTCACAGGATGGTGACTTGTTATGGAGTATGGCTGCTGCGTCA
AAATGTTTGATGAAGAAGGAACCTGAAGTTATCATGAGACCGTTAGTATGGG
ATGAATTCGCCAACAATACAGATTGCGACAAATCGGTAATTTTACTTTAATAT
GCTAATAGGGAATTGAACTAATGTTTTCCAAGCGTTTGCAGGTATTCGCGTG
TCATGCATCGAATTCTTACATCAATGGTGTGCGGAGCAAGCTTCGAAACACG
AATAGCGCCGACGCCAACTCAGAAAGGACGATTGTGTGTCGCTGATCAGT
CGAGCCAAGGATTCGGATAATCTGGCCCGAATGCCACCCACCTACCACGC
GTGGTTCTAGATCTCATAATTACCGTTCTCTATTTTGATCCCGCCTCCCCTC
TCACAGATCTCTATACATTTGTCAAATGTTTCCAAATCTTTTATCTGCCATA
CATTGTTTTTATTGTTTTGTTTCTTTTCTTTCTTTATTTCTTTTCTAACTTTA
AGATTTATGTAAATATTTAACTGCGCTGGTATTTATGAAAAATTCAGATAAAG
TTTTCAAGTTTAAAAAATCGAAAATTCGAAGTCGGAAGTTCTCTTACAGGTGT
AGTAAGTAGGCACAATGGCAATAGGTACATGGAAGGCTTGCGGAAGGCACA
TGGGTAGGCATAAGATCGAAAAATAAGCTGATATATAAATATAGATAGGTAT
TGTTAGGCACAAATTAGGCACGTAGGTGTGAGCTGGCAAATAGGTAGGCA
TGACGTTTCGGCAAATCGGCAAATTGCCGATTTGGCGAAAATTTCAAATCCG
GCGATTTGCCGGAAATGTTTAGAGAAATTTTTTATAAGACAGAAAAACTTACA
ACTGTGTCTTTTTGAAATCTTCCGGTTTTCTTTATACAGTGCGTGCAACTTC
TATAGCGCCCCCCCCCCCCCCCCCCCCCTATTTTTTGGCGTTTCACGCC
ATTCTGATTTTTATTTTTCTGATTTTTTTTTTTTTTGCCTGAACTTGGCATTTGA
GGATGCTTGGAGAGAAATATCAGCCAGCAAATAAAGAATCTGGTCAACTCA
ATGTCGAATAGATTTTTGAGGTTATCGTTAAGAAGGGAGGTCCCACGACGT
ATTGATCCTTCATCGAGTTAACAATTATGATGTTTTAATTGATTTCAATCCAC
TTCTGGACACAGAAGGACGAATAGTGCAATCTGGTACAAGTTTATCACCACC
TACAACTTCGTCGATTTGTGGAAAATCTTTCAGACATGTCTCCATGAGTGTC
TCAGAACATCTTGGTCAGGTTTGGAGTCGATCCCACCGCTGGGAGCCGAGA
ATGGGCCTCTAACAC

FIGURE 9

itr-1 ORF sequence

ATGGATCCGGCTATGGCTTCTCCAGGCTATCGGTCTGTGCAGTCCGATCGG
AGTAATCACCTAACAGAGCTGGAAACGAGAATTCAAATCTTGCCGATAATT
CACAAAGAGATGATGTCAAATTGAAATGTTACAAGAGATTGGAGCACAAAT
CGAAAATCATTTCACACTAAGTTCGCACGAGAAAGTCGTGGAGAGGCTCATT
CTCTCGTTCCTACAAGTTTTCTGCAACACAAGTCCACAGTTCATTGCTGAAA
ACAATACACAACAGCTTCGAAAGTTAATGCTTGAAATCATTCTTCGACTTTTCG
AACGTAGAAGCCATGAAACATCATAGCAAAGAAATTATCAAGCAGATGATGA
GGCTAATCACCGTGGAAAATGAGGAGAATGCCAATTTGGCTATCAAATTTGT
CACCGATCAAGGGAGAAGTACCGGCAAAATGCAATATTGCGGAGAGGTTTC
ACAGATAATGGTCTCCTTCAAACAATGGTCATTGATCTGACGGCGAGTGGT
CGAGCTGGTGATATGTTCAACATAAAAGAGCATAAAGCTCCACCGTCAACTA
GCTCCGACGAGCAAGTCATCACTGAATATTTGAAGACTTGCTACTATCAACA
AACGGTTCTTCTCAACGGAACGGAAGGAAAACCGCCATTAAATACAATATG
ATTCCATCAGCTCATCAGTCAACGAAGGTGCTCCTGGAGGTTCCGTATCTC
GTGATTTTCTTCTATCAACATTTCAAACAGCGATCCAAACCGAAGCGCTTG
ATTTTCATGAGGCTTGGTCTTGATTTTCTAAATGTCAGAGTTCAGACGAGGA
TAAACTCAAACAAATCAAATAATAACCGATGATTTTGTGAGTGCACAGTCCC
GATTCCTGTCAATTCGTCAACATTATGGCTAAGATTCCAGCGTTTATGGATCTT
ATCATGCAAATGGACCGCTTCTAGTGTCCGGAACAATGCAGATGCTCGAG
CGGTGCCCGGCTGATCTGATAAGTGTCCGACGAGAAGTTCTGATGGCTTTG
AAGTATTTACATCTGGAGAAATGAAGTCGAAATTTCTTCCAATGCTACCTC
GACTCATCGCTGAGGAGGTTGTTGTGGGAACAGGATTCATGCGATTGAGC
ATTTGCGAGTTTTTCATGTATCAAATGCTAGCAGATCTGTTGCATCACATGCG
AAATTCTATAGACTATGAAATGATCACACACGTGATTTTCGTATTCTGTGCA
CTCTTCACGATCCTAACAACCTCTTCTCAAGTCCAGATTATGTCTGCTCGGCT
GCTCAACTCACTGGCCGAATCTCTGTGCAAAATGGATTCACATGATACCTTT
CAGACTCGTGATCTGCTCATTGAAATCCTGGAGTCGCACGTGGCCAAAGCTC
AAAACCTTTGCAGTCTATCACATGCCTATTCTCTTCCAACAATACGGAACCG
AAATAGACTACGAATACAAAAGTTATGAGAGAGACGCCGAGAAACCTGGAA
TGAATATCCCAAAGGACACTATACGAGGAGTACCGAAACGAAGAATCCGTC
GGCTCTCCATTGATTCAGTTGAAGAGCTGGAATTCCTGGCATCAGAACCATC
CACGTCCGAAGATGCAGATGAGAGTGGTGGAGATCCGAACAAGCTTCCTCC
GCCAACAAAAGAGGGGAAAGAAAACGTCTCCCGAAGCGATTTTAACCGCCAT
GTCAACGATGACACCTCCTCCATTGGCAATTGTTGAAGCTCGAAATCTTG
AAGTATATAATGCATACGTGTAAATTCGTGACAGGACAATTGAGAATCGCCC
GGCCATCACAGGATATGTATCATTGTTGGAAGGAGCGAGATTTATTGGAACG
TCTTCTACGATATGGTGTAAATGTGTATGGATGTATTGCTGCTTCCAACAAC
CGAAATCAACCACAAATGCATTCTTCAATGCGGACAAAAGATGAGAAAGATG
CTCTGGAGTCTGTTGGCAAACGTTTTTACAACAATCGACCATGCGATATTCCG
GGAAATCTTCGAAAAGTATATGGATTTCTTGATTGAAAGAATTTACAATCGGA
ACTATCCATTGCAATTGATGGTGAACACCTTCTTGGTTCGAAATGAAGTGCC
ATTCTTCGCATCTACGATGCTTTCATTCTTGATGTCTCGAATGAAATTGCTGG
AAGTTAGCAATGACAAGACGATGCTATATGTGAAGCTCTTCAAATATCTTC
TCCGCCATCGGAGCCAATGGCTCTGGGCTTCATGGAGATAAAATGCTCACT
TCATACCTCCCAGAGATTCTCAAACAGTCAACTGTCTTGGCATTAAACAGCTC

19/92

FIGURE 9

GTGAACCTCTCAACTATTTCTTTTGCTTCGTGCATTGTTCCGCAGTATTGGT
GGTGGCGCTCAGGATAATTTGTATGGAAAGTTCCTGCAGTTACTGCCAAATC
TTCTTCAATTCTTGAATAAATTGACGAATCTTCAGTCATGTCAACATCGGATT
CAAATGCGTGAGCTCTTCGTGAGTTGTGTTTGAAGTGTGCCAGTTGCACTCA
GTTCCCTTCTGCCATACCTACCGCTTCTGATGGATCCACTGGTGTGTGCGAT
GAATGGGAGTCCGAACATAGTTACACAAGGATTGAGAACATTGGAATTATGT
GTGGATAACTTGCAACCTGAATATCTTCTCGAAAATATGCTTCCTGTCCGTG
GAGCTTTGATGCAAGGCCTCTGGCGTGTGATCGAAAGCTCCAGATACAT
CATCGATGACAGCAGCGTTCAGGATCCTCGGAAAGTTCGGAGGAGCCAATC
GAAAACCTTCTGAATCAACCGCAAATTCTTCAAGTAGCCACTTTAGGCGACAC
TGTTCAAGTCGTACATCAATATGGAATTCTCGCGGATGGGACTCGATGGCAAT
CACAGCATTACCTGCCACTGTCCGAGTTGATGAGAGTCGTTGCCGATCAG
ATGAGATATCCAGCTGATATGATCCTTAATCCAAGTCTGCAATGATCCCGT
CAACTCATATGAAGAAATGGTGTATGGAATTGTGAAAGCCGTCTTGTTAGC
CGGACTTGGATCTTCAGGAAGCCCAATTACTCCAAGTGCAAATCTTCCGAA
GATTATCAAGAACTTCTTGAAGATTTTGATCCAAACAATCGTACCACTGAAG
TATACACATGTCCGAGGGAAAGTGATCGAGAGCTTTTGTGAATGCACTTCT
CGCAATGGCTTACGGAATATGGAATAAAGACGGTTTCCGGCATGTCTATAG
CAAATTCTTTATCAAAGTTCTCCGCCAGTTTGGCTTGATTGGAGTACTCGAA
TACATTGGTGGAAATGGATGGATGCGTCATGCAGAAGAGGAAGGTGTTCTA
CCATTGTGCCTTGACTCGTCTGTTATGGTTGATGCTCTGATTATTTGTCTCTC
TGAAACATCGTCAAGCTTCATCATTGCTGGTGTGATGTCTCTTCGTCATATC
AATGAGACTCTCTCGCTTACACTTCCCGATATTGATCAAATGTGAAAGTTC
CAATGTGCAAATACTTGATGGAGAAGGTGTTCAAATTGTGTACGGGCGCTG
CTTGGTATGCAAGATCTGGTGGAAATCAATGCAATTGGATACATGATCGAATC
GTTTCCACGAAAATTTGTTATGGACTTTGTGATAGATGTTGTTGATTGATCA
TGGAAGTTATTTTGGGAACTGTTGAAGAAATATCAAGTGGATCTGCTGATTC
TGCATACGATTGTCTCAAGAAATGATGCGAGTCTATTTTATCAAAGAAGAA
GGCCAAGAAGAGGAGAATCTGACACTCGCGACTATTTTGTGTCTGCAATCT
CTAAGCATTACTTCCACAGTAATGAAAGAGTCAGAGAATTTGCGATTGGTTT
AATGGATCATTGTATGGTTCACTCAAGACTTGCACCATCCCTTGATAAGTTC
TACTATCGATTCAAGGAGTTCTTTGAGCCAGAATTAATGCGGGTGCTCACAA
CAGTTCCAACAATGTCATTGGCAGACGCAGGAGGAAGTTTGGATGGAGTTC
AAAACATATGTTCAACTGTCCGGATGGTTTTGATTTGAAAAAGATATGGA
CATGTACAAGCGATATTTGTCACATCTGCTGGATATTGCACAAACCGATACA
TTTACCTTAAACCAAAGGAATGCCTTCAAAAAATGCGAGACATGCCCATCGC
ATTTCTTCTCCTCCATTCCAATCACTACACATATTGATTCAATGCGAGCCAGT
GCTCTACAGTGTCTTGATGATCGCGTATGATCGAATGAAGAAGCAATACATCG
ACAAGGGAATAGAGCTGGGTGATGAGCATAAGATGATAGAGATCCTCGCAC
TTCCGAGCTCCAAGATCACAGTTGATCAAGTCTACGAGAGCGATGAATCTTG
GAGACGATTGATGACAGTTCTATTGAGAGCAGTCACTGACAGAGAACTCC
TGAAATTGCGGAGAAGCTTCATCCTTCACTTTTGAAGGTCTCACCAATATCC
ACAATCATCATCGCAACATTTGGTGCTTCTTACATAAGAAATATTAGTGGAG
CAGGAGATGACAGTGATTGAGATCGTCATATTTGATACAACGATATAATGAA
GTTCAAGTGTCTCGTGGAGCTCAATCCAAGATTCTGGTCACAAAAATGGCA
GTGAATCTCGCAAATCAAATGGTTAAATATAAGATGAGTGACAAGATCTCTA

20/92
FIGURE 9

GGATTTTGTGAGTTCCAGTAGCTTCACTGAAGAGGAGCTCGATGATTTGGA
AGCGGAGAAGATGAAAGGAATTCGAGAGTTGGATATGATTGGTCATACGGT
TAAAATGCTTGCTGGATGCCAGTGACCACATTCACGGAGCAAATTATTGTG
GATATCAGTCGTTTGTGCTGCTCATTTTGAGTATGCTTATTCGCAAGATGTACT
TGTAATTTGGATTGATGATGTCACAGTAATCCTCAACAAAAGTCCCAAAGAT
GTATGGAAGTTCTTCTTGTCTCGAGAATCAATTCTAGATCCTGCACGCAGAT
CCTTTATTGGAAGAATCATAGTCTATCAATCAAGTGGTCCACTGCGACAGGA
ATTCATGGATACTCCGGAATATTTGAGAACTCATTGATCTTGACGATGAG
GAGAATAAGGATGAAGATGAGAGAAAAATCTGGGATCGTGATATGTTTGCAT
TTTCGATTGTGCGATCGTATCTCGAAGAGCTGCCCTGAGTGGCTTATTTCTCC
GAATTCCTCAATTCCAAGAATTAAGAAGTTGTTCTCCGAAACGGAATTCAT
GAGCGATATGTGGTTCGAGCATTGACTGAGGTGAAGAAATTTCAAGAAGAG
ATCATAGTGAAACGGATGACAGAGCACAAGTACAAGGTTCCGAAGCTGATT
CTGAATACCTTCTGAGATATTTGAGGCTCAACATCTATGACTACGATCTATT
CATCGTTATCGCCTCGTGTTCATGGCAATTCGTACCGGATCTCTCTTTTC
TTCGCGAATATCTTGAACTGAAGTCATCCCGAAAGTGCCGTTACAATGGCG
GAGAGAGCTGTTTCTTGAATTATGCAGAAGTTTGATACGGATCCACAACT
GCTGGAACAAGTATGCAGCATGTGAAGGCCCTTCAATATTTGGTTATTCCCA
CGTTGCATTGGGCGTTGAGCGATATGATACGGATGAAATTGTTGGCACC
CACCAATAGATGATTGCGATTCTTCGATGGATGTAGATCCGGCAGGCAGCT
CGGATAACCTTGTGGCTCGTTTAACATCAGTCATTGATTCTCATCGTAATTAT
CTGAGCGATGGAATGGTCATTGTTTTCTATCAACTTTGCACATTGTTGCTAC
AAAACGCCTCCGAACATATTCACAATAATAACTGCAAGAAACAAGGTGGACG
CCTACGGATCCTGATGCTCTTCGCCTGGCCGTGCCTGACCATGTACAATCA
TCAAGATCCAACAATGEGGTAGACTGGATTCTTCTTCTTGGCCAATATTATA
GAGCGTTTCACAATTAATCGGAAAATCGTGCTTCAAGTGTTCCATCAACTTA
TGACTACTTATCAGCAGGACACTAGAGATCAAATCCGGAAAGCCATTGATAT
ATTAAGTCCAGCTTTGAGGACACGAATGGAAGATGGACACTTGCAAATATTG
AGTCATGTGAAGAAAATTCTTATCGAAGAATGCCATAATTTGCAACATGTTCA
GCATGTTTTCCAAATGGTGGTTCGCAATTATCGTGTCTACTATCATGTTGAT
TGGAGCTTCTCACGCCTCTTCTGAACGGAGTTCAACGAGCACTTGTGATGC
CAAATAGTGTTCTGGAAAAATTTAGCTGGCAAACCTCGACGTCATGCGGTGG
AGATCTGCGAGATGGTCATCAAGTGGGAATTGTTGAGAAGCTGAAAACAG
ATCATATTATCAGTGACGAAGAAGCTCTCGAAGTTGACAAGCAATTGGATAA
GCTGCGAACAGCTTCATCCACAGATCGTTTCGATTTCGAGGAGGCTCATAA
CAAGAGAGACATGCCTGATGCTCAACGCACGATTATCAAAGAGCACGCCGA
TGTGATTGTCAATATGCTTGTCCGATTCTGTATGACGTTCCATCAGAATTG
GGTCTTCGTCCACTTCTCAAAGTGGGAACCATGGTGTGAGTTGACCAAA
AATGTGAGCTGCTTCTACGTGCAGCCCTACGACCAAGCATGTGGGGAGAA
TTTGTGAGCTTCCGATTAACAATGATCGAAAAGTTTTTGTCAATTCCGAATGA
TAATGCTCTACGCAATGATATAAGTTCTACGGCCTACGCTAATACTATCCAA
AATGCAACAACACTCTGGATATGCTGTGTAATATTATTCTGTTATGCCAAA
AACTAGCTTGATGACTATGATGAGACAACCTCCAACGGCCACTCATACAATGT
CTCAATAACGGAGCTCAGAACTTTAAGATGACTCGTCTTGTCACTCAAATTG
TCAGTCGGTTACTCGAAAAGACAAATGTTTCGGTTAACGGGCTTGATGAGCT
GGAGCAATTGAATCAATACATTTCCCGATTCTACATGAACATTTTGGATCTC

21/92

FIGURE 9

TTTTGAATTGCAGAACTTGAGTGGACCAGTGTTGGGAGTTCTCGGAGCATT
TTCTCTTTTGC GAACAATTTGTGGACACGAGCCAGCATACTTGGATCATTG
ATGCCCTTCATTTGTAAAAGTGATGGAGAGAGCTGCAAAAGAGCACTTGGCG
TATGTTGCGAACTCGCAAGATGGAAATATGGTGAAGAATTTCTTTCCAGATG
TTGCTGAATTGTTGTGTGCATGCATGGAGCTGGTACGTCCCAGAGTCGATC
ATATCAGTATGGAGATTAAGAGATCAATTGTTGGTGGTATTATCGCGGAGCT
GATTATCAAATCGAATCACGATAAGATCATCCAGACGTCAGTGAAGCTTCTC
GGAGCAATGATTAGCACGCAGGATATGGAATTTACAATTCTCACTGTTCTTC
CGCTACTTGTTTCGTATCCAATCAATTATTGTGACCAAGTTCAAGAATTGCAA
GGATCTGATAGCAGACTATCTTGTTGTGGTTATTACCGTTTTTGAGAACAGC
GAATATCGGAACCTCGGAAGCTGGATCTCGTCTCTGGGAAGGATTCTTCTGG
GGACTCAAGAGTAGCGATCCTCAAACCCGGGAGAAATTCGATAGTTTGG
GAGAAGACTTGGCCACACATGGCAACAGTAGATATTGCTCATCGAATGAAAT
ATATCATGCAAAATCAAGATTGGTCCAAGTTCAAACACGCGTTTTGGTTGAA
ATTCGCACTTTGGGGAATGCTACGAACGATTGCCAAACGGCCAACCTGATCC
GAATAATAAGAGAAAGAAAGTGATACTGTTGAACTGTGCAACTCCATGGAGA
ACAATTGAATATGCAGCGAAATTGAAGGATCAGCCAATGGAAGTGGAACT
GAAATGAAACGAGAAGAGCCAGAACCGATGGAAGTTGACGAAAAAGACTCG
CAAGATGATTCTAAGGATGCCGGAGAGCCCAAGGAGAAGGAAAAGCTCACA
TTGGAATTATTGCTTGCTGGACAACAAGAACTTTTTGGATGAAGCTTCCAATT
ATGATTTTGCGGATGCTCTAGATACAGTATCCCAGATTACATTTGCACTTAAT
GAGAATCAAGTGACAAGCAAGATGTGGGTAGTGTGTTGTTCAAATCATTCTGGA
GTTCTTATCACAATCCGAAATCGAAGATTTACGGCGCTAGTCGTTCCGTT
TATGAGCAGTGGAGTGCATAATAATTATCAGACGGGTGTACAGGATAGTGT
GCTTGCTGTTTGGCTTGAAGCTGTTGGTGACGCTGTTCAATTTGCCGTCCAG
ATTGATTGAGTTTATCTCATCAAAACACGAATGCTGGCATACCGGAATCAGG
CTTCTCGAGAATCATATATGGACAATTCAAAGCAACTCAACAACACGTTAC
TCCGAGAAATGAAAGTGGCACCAGGTCTCGCTGGAGATATTGAGACACTCG
AATCTCTTGGAACACTCTACAATGAGATATCAGAGTTTGATCAGTTCGCTGC
AATCTGGGAACGCCGTGCTGTATTTCTTGATACGATGAGAGCAATGTCAGC
TATGCAATTGGGAGATATGGAATTAGCTCAATCTTATCTGGAAAAATCAATG
AGCAGTACGTATGAAACTCTTGCTCCGACAATCAATCCAAACAACACTTCAA
ATTCCGAGAAGCATGTTTCTCCGATTATTGACAAAGAATACGATCATTGGAT
GGAGATGTACATCACAAATTGCTCGGAGCTTCTTCAGTGGCAAAATGTGGC
CGACGTATGCAATGGCAAAGACATGCAACATGTTCTGTTGGCCTGATCAACGC
AGCATCTCACATTCCGGACTGGAATGTGGTCGAGGAGTGTAAGAGTCAGAT
AGCTGGATGTATTCCACCAAGTTTCCATTTAGATTACACTCTTTTCAATTTGA
TGAGTACTGTTATGCGAATGAATGAAAACCTCAAGCCCGACACATATGAAGGA
ACGATGCAAAATTGCAATTCAAGAGTGCACAGAAGCTCATATTAGTCGTTGG
AGAGCACTTCCGTCAGTTGTTTCATATGGTTCATGTCAAGATTCTTCAGGCAA
TGAACCTGGTTTCGAGAAATTGAAGAGTCTACAGATATTGCGATTGCTCTGCT
CGAGGCCCATCAAACAAAGTGGATCAGGCGTTGATGGGCGATATGAAGTC
GTTGATGAAAGTATTCCGAAATAGAACACCAACCCTTCGGATGATATGGGA
TTCGTTTCGACTTGGTATGATTGGAGGAATCAGATTCATGGAATGATGCTTC
AAAGATTCGAATATTGGGATAAAGTAGGACTCAACGTCGCTGCAACTGGAAA
CCAGTCAATTGTTCCGATTCATTCAATGGCTCAAGCACAGTTGGCCGTAGCC

22/92
FIGURE 9

AAACATGCCAAGAATCTTGGATTCCATAATTTAACGAAAGATCTACTCAACAA
ATTAGCTGGATTGACAGCCATACCGATGATGGATGCTCAAGATAAAGTTTGC
ACTTACGGCAAGACACTTCGCGATATGGCAAACAGTGCGGCTGACGAAAGA
GTGAAAAATGAGCTATTGTGTGAAGCGCTTGAAGTTTTGGAAGATGTGCGAA
TTGATGATCTACAGAAGGATCAGGTTGCTGCATTGCTTTATCATCGTGCTAA
TATTCATTCAAGTTCTTGATCAAGCTGAAAATGCTGACTACACCTTCTCCGCA
GCCTCTCAACTTGTGCACTTGCAAATAGTGTGACAACCACTGGAATCAAGC
TCATGAAAAATTGGGGCCACCATCTTTACAAGAGATTCTTCTCTACGACAGT
TTGCAAGGAAACCGGAAACAACCTTCGGACGGCAGGCTCTCGCTTGTTACTT
CATTGCGGCTCGTGTGGATAACGATATCAAGGCGAGAAAACCGATTGCCAA
GATTTTGTGGCTCTCGAAGCACTTGAATGCGTGTGGATCACATGAAGTGAT
GAATCGGGTTATTAAGAAGCAACTTCATTCACTTAATCTCTTCAATTGGCTTT
ACTGGCTTCCACAATTGGTTACTGATGTTGATATAAACCAAATTCGAACCTT
GTTCTGATTCTCTGCAAGATGGCTGCTGCTCATCCACTTCAAGTATTTTACC
ACATTCGGGAGGCAGTTAGCGTTGACGATATTGACTCGGTTCTCGAAGAAG
ATTACACTGATGAGCAAATGTGATGGATGTTTCGGATGAGGATTGTTTTGC
AGACGATCCACCATTTGATAGAATTCTGAAAATATGTCTGAAATATCGTCCAA
CTGATATTCGAGTCTTCCATCGTGTCTCAAAGAACTTGACGAGATGAATGA
GACATGGGTTGAACGTCACTTGCGTCAATGCGATCTGCCTCAAGGATCAGAT
GTTCAAAGATTTCTCGGAACAAATGGACGCGACGTTCAATGAGATGCAATAT
TCGGAGGATGTGACTATGATGACGTTGAGATGGAGGAAACAGCTGGAAGAA
GACTTGGTGTATTTCCAACAGAATTATAATCTTGATTTCTTGAGATTGCTAA
CAAGCGAAAGATGATCGTGACGAAGGGATGTATGGGAGTCGAGAAAAGTCA
GATAATGTTGCAAAAAGAGCTGAGTCAAGTGTTACAGAGCCGGCCGGCAT
GCAAGATGAATTTGATTTTGTACAAATATGACTAATATGATGGTCTCAGAT
TGGATATTCATGCAGTCGATGCTCCACGCCCTCAGGGATATATTCGTATTGT
TCTCGACTGGATTGAGCGATTGCTGCTGCTTTCGATCGACTTCCACGAAG
AATCCCTCTGGAATCGTCAAGCCCATATCTCGCCAGATTGAGCCATCGTACA
GGATGCATCGAAATGCGATAGGATTTGCTCAACGTTTTGCGCGCCAAGAAT
CATACTCTGATGGCTTCCAATCAAACGGGGCAATACATATCCATGCTCTCTC
GATTTGAGCCAACTTTGAGATTGTGATCAAAGGTGGTCAAGTGATAAGAAA
GATCTATATTCGAGGACAAACCGGAAAGAGTGCGGCGTTTTATCTGAAGAA
ATCTGTGAGGATGAGCCAACTAACCAGATTCCACAAATGTTCAAACATCTT
GATCACGTTCTACAAACCGATAGAGAGTCGGCGAGAAGACATCTTCATGCT
CCAACAGTGCTGCAGATGAGAGTCGGACAGAAGACGACACTCTACGAAGTT
GCATCCGTTCAACCATATGCAATGCCACCGGATTGTACCAGAACTATCCAG
CATCACAAATCGACATTGTTCCATCATATGATGTGCTGACTGCCACTTTCAAT
GGAAGTTATTATCCGGATGATATGGTATTGCACTTCTTTGAGAGATTGCGCC
AAAGTTCTTCATCCATCGGACAACCTCTTCCAACCTCCGACGAACCAAGATGG
AACAGTTGCTCCGCCACGACTAACGGAAGCTCACCACATCAAGAATATTATT
TATGAAGACTTTGCCCGAGATATGATCCCATTCGACTTCTCTACGACTACC
TCACTGCACGATATCCTGATCCGGTTATGTACTATGCAATGAAGAAGCAATT
GCTGCACAGTCTCGCGGTGCTATCCACAATCGAATATCATTGCAATCTGA
CCAATGGGACCTGATCAAATGATGATGACAATGAATACTGGAGTCCTTAGCA
ATCCTTCATATAGATTGCAAATCCGAGGAGGACGATCACTTCATGATATTCA
ACACTTTGGACATGAAGTTCCATTCCGATTGACTCCAAATCTATCGATTTTG

FIGURE 9

GTTGGTGTTCACAGGATGGTGACTTGTTATGGAGTATGGCTGCTGCGTCA
AAATGTTTGATGAAGAAGGAACCTGAAGTTATCATGAGACCGTTAGTATGGG
ATGAATTCGCCAACAATACAGATTGCGACAAATCGCGTTTGCAGGTATTCGC
GTGTCATGCATCGAATTCTTACATCAATGGTGTGCGGAGCAAGCTTCGAAAC
ACGAATAGCGCCGACGCCAAACTCAGAAAGGACGATTGTGTGTCGCTGATC
AGTCGAGCCAAGGATTCGGATAATCTGGCCCGAATGCCACCCACCTACCAC
GCGTGGTTCTAG

24/92

FIGURE 10

TRR-1 protein sequence

MDPAMASPGYRSVQSDRSNHLTELETRIQNADNSQRDDVCLKMLQEIWSTIE
NHFTLSSHEKVVERLILSFLQVFCNTSPQFIAENNTQQLRKLMLEIILRLSNVEAM
KHHSKEIKQMMRLITVENEENANLAIKIVTDQGRSTGKMQYCGEVSQIMVSFKT
MVIDLTASGRAGDMFNIKEHKAPPSTSSDEQVITEYLKTCYYQQTVLLNGTEGK
PPLKYNMIPSAHQSTKVLLVPYLVIFFYQHFKTAIQTEALDFMRLGLDFLNVRV
PDEDKLKTNQIITDDFVSAQSRFLSFVNIMAKIPAFMDLIMQNGPPLLVS GMTMQL
ERCPADLISVRREVLMAKYFTSGEMKSKFFPMLPRLIAEEVVLGTGFTAIEHLR
VFMYQMLADLLHMRNSIDYEMITHVIFVFCRTLHDPNNSSQVQIMSARLLNSL
AESLCKMDSHDTFQTRDLLIEILESHVAKLKTAVYHMPILFQQYGTEIDYEYKSY
ERDAEKPGMNIPKDTIRGVPKRRIRRLSIDSVEELEFLASEPSTSEDADESGGDP
NKLPPPTKEGKTSPEAILTAMSTMTPPPLAIVEARNLVKYIMHTCKFVTGQLRIA
RPSQDMYHCSKERDLFERLLRYGVMCMDFVLPTTRNQPMHSSMRTKDEK
DAESLANVFTTIDHAIFREIFEKYMDFLIERIYNRNYPLQLMVNTFLVRNEVPFF
ASTMLSFLMSRMKLLEVSNDKTMLYVKLFKIIFFSAIGANGSGLHGDKMLTSYLPE
ILKQSTVLALTAREPLNYFLLLRALFRSIGGGAQDILY.GKFLQLLPNLLQFLNKLT
NLQSCQHRIQMRELFVELCLTPVVRSSLLPYLPLLMDPLVCAMNGSPNIVTQG
LRTLELCVDNLQPEYLLENMLPVRGALMQGLWRVVS KAPDTSSMTAAFRILGK
FGGANRKL LNQPQILQVATLGDTVQSYINMEFSRMGLDGNHSIHLPLSELMRVV
ADQMRYPADMILNPSPAMIPSTHMKKWCMELSKAVLLAGLGSSGSPITPSANL
PKIIKKLLEDFDPNNRTTEVYTCPRES DREL FVNALLAMAYGIWNKDGFRHVYS
KFFIKVLRQFALIGVLEYIGGNGWMRHAE EEGVLPLCLDSSVMVDALIICLSETS
SSFIIAGVMSLRHINETLSLTLPDIDQMSKVPMCKYLMEKVFKLCHGPAWYARS
GGINAIGYMIESFPRKFVMD FVIDVDSIMEVILGTVEEISSGSADSAYDCLKKM
MRVYFIKEEGQEEENLTATIFVSAISKHYFHSNERVREFAIGLMDHCMVHSRLA
PSLDKFYYRFKEFFPELMRVLT TVPTMSLADAGGSLDGVQNYMFNCPDGFDF
EKDMDMYKRYLSHLLDIAQTD TFTLNQRNAFKKCETCPSHFLPPFPITTHIDSMR
ASALQCLVIAYDRMKKQYIDKGIELGDEHKMIEILALRSSKITVDQVYESDESWR
RLMTVLLRAVTDRETPEIAEKLHPSLLKVSPISTIIATFGASYIRNISGAGDDSDS
DRHISYNDIMKFKCLVELNPKILVT KMAVNLANQMVKYKMSDKISRILSVPSST
EEELDDFEAEKMKGIRELDMIGHTVKMLAGCPVTTFTTEQIIVDISRFAAHFEYAY
SQDVLVNWIDDTVILNKSPKD VWKFFLSRESILDPARRSFIRRIIVYQSSGPLRQ
EFMDTPEYFEKLIDL DDEENKDEDERKIWD RDMFAFSIVDRISKSCPEWLISPNS
PIPRIKKLFSETEFNERYVVRALTEVKKFQEEIIVKRMTEHKYKVPKLILNTFLRYL
RLNIYDYDLFIVIASCFNGNFVTDL SFLREYLETEVIPKVPLQWRRELFLRIMQKF
DTPQTAGTSMQHV KALQYLVIPTLHWA FERYDTDEIVGTAPIDDS DSSMDVDP
AGSSDNLVARLTSVIDSHRNYLSDGMVIVFYQLCTLFVQNA SEHIHNNNCKKQG
GRLRILMLFAWPCLTMYNHQDPTMRYTGFFFLANIIERFTINRKIVLQVFHQ LMT
TYQODTRDQIRKAIDILTPALRTRMEDGHLQILSHVKKILIEECHNLQHVQHVFQ
MVVRNYRVYYHVRLELLTPLLNGVQRALVMPNSVLEKFSWQTRRHAVEICEMV
IKWELFRTLKTDHII SDEEAEV DKLRTASSTDRFD FEEAHNKRDM PDAQ
RTIIKEHADVIVNMLVRFCMTFHQNSGSSSTSQSGNHGVELTKKCQLLLRAALR
PSMWGEFVSFRLTMIEKFLSIPNDNALRNDISSTAYANTIQNAQHTLDMLCNII PV
MPKTS LMTMMRQLQRPLIQCLNNGAQNFKMTRLVTQIVSRLL EKT NVS VNGLD
ELEQLNQYISRFLHEHFGSLLNCRNL SGPVLGVLGAFSLLRTICGHEPAYLDHL
MPSFVKVMERAAKEHLAYVANSQDGNMVKNFFPDVAELLCACMELVRPRVDHI

SMEIKRSIVGGIIAELIISNHDKIIQTSVKLLGAMISTQDMFTILT VLP LLVRIQSII
VTKFKNCKDLIADYLVVVITVFENSEYRNSEAGSRLWEGFFWGLKSSDPQTREK
FSIVWEKTWPHMATVDIAHRMKYIMQNQDWSKFKHAFWLKFALWGMLRTIAKR
PTDPNNKRKKVILLNCATPWRTIEYAAKLKQDPMEVETEMKREEPEPMEVDEK
DSQDDSKDAGEPKEKEKLTLELLLAGQQELLDEASNYDFADALD TVSQITFALN
ENQVT SKMWWVLFKSFWSLSQSEIEDFTALVVPFMSSGVHNNYQTGVQDSV
LAWWLEAVGDAVHLP SRLIEFISSKHECWHTGIRLLENHIWTIPKQLNNTLLREM
KVAPGLAGDIETLES LGTLYNEISEFDQFAAIWERRAVFPDTMRAMSAMQLGD
MELAQSYLEKSMSSSTYETLAPTINPNNTSNSEKHSVPIIDKEYDHWMEMYITNC
SELLQWQNVADV CNGKDMQHVRGLINAASHIPDWNVVEECKS QIAGCIPPSFH
LDYTLFNL MSTVMRMNENSSPTHMKERCKIAIQECTEAHISRWRALPSVVSYG
HVKILQAMNLVREIEESTDIRIALLEAPSNKVDQALMGDMKSLMKVFRNRTPTTS
DDMGFVSTWYDWRNQIHGMMLQRFEYWDKVG LNVAAATGNQSIVPIH SMAQA
QLAVAKHAKNLGFHNLTKDLLNKLGLTAIPMMDAQDKVCTY GKTLRDMANSA
ADERVKNELLCEALEVLEDVRIDDLQKDQVAALLYHRANIHSVLDQAENADYTF
SAASQLVDLQNSVTTTGIKLMKNWGHLYKRFFSTTVCKETGNNFGRQALACY
FIAARVDNDIKARKPIAKILWLSKHLNACGSHEVMNRVIKKQLHSLNLFNWLYWL
PQLVTDVRYKPNSNFVLILCKMAAAHPLQVFYHIREAVSVDDIDSVLEEDYTDEQ
MSMDVSD EDCFADDPPFDRILKICLK YRPTDIRVFHRVLKELDEMNETWVERHL
RHAICLKDQMFKDFSEQMDATFNEMQYSEDVTMMTLRWRKQLEEDLVYFQQN
YNLDFLEIRNKRKMIVTKGCMGVEKSQIMFEKELSQVFTEPAGMQDEFDFVTN
MTNMMVSQLDIHAVDAPRPOGYIRIVLDWIRAIRRRFDRLPRRIPLESSSPYLAR
FSHRTGCIEMPYDLLNVLRANKNHTLMASNQTGQYISMLSRFEPNFEIVIKGGQVI
RKIYIRGQTGKSAAFY LKKS VQDEPTNRVPOMFKHLDHVLQTDRESARRHLHA
PTVLQMRVGQKTTLYEVASVQPYAMPDPCTRNY PASQIDIVHPYDVL TATFNG
SYYPDDMVLHFFERFAQSSSSIGQPLPTPTNQDGT VAPPRLTEAHHIKNIYEDF
ARDMIPFRLLYDYLTARYPD PVMYYAMKKQLLHSLAVLSTIEYHCNLT PMGPDQ
MMMTMNTGVLSNP SYRFEIRGGRSLHDIQHF GHEVPFRLTPNLSILVGVAQDG
DLLWSMAAASKCLMKKEPEVIMRPLVWDEFANNTDCDKSRLQVFACHASNSYI
NGVASKLRNTNSADAKLRKDDCVSLISRAKSDNLARMPPTYHAWF

FIGURE 11

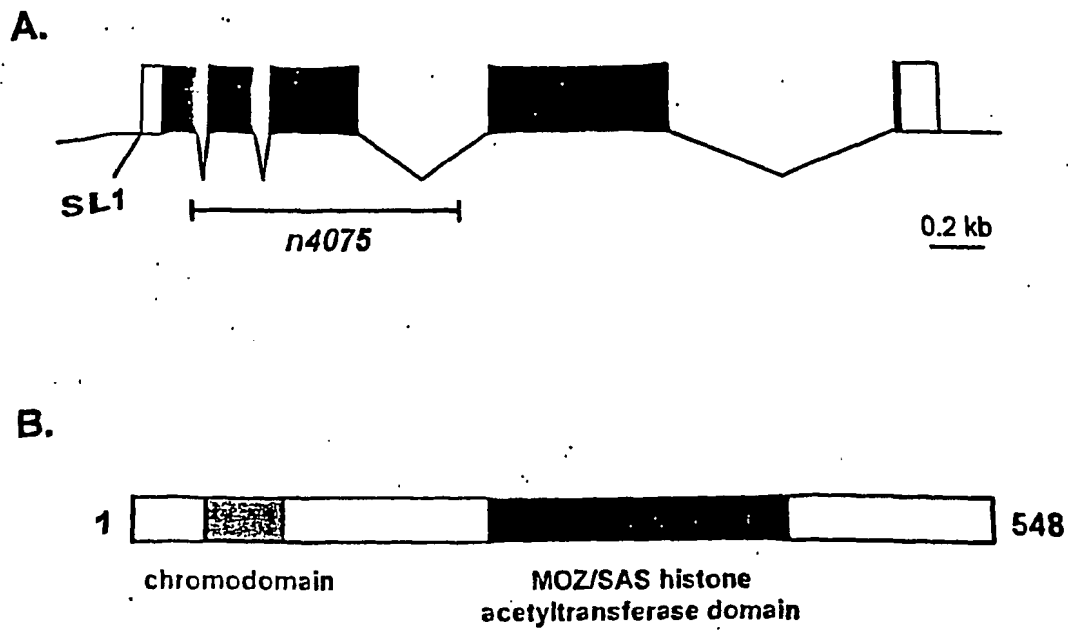
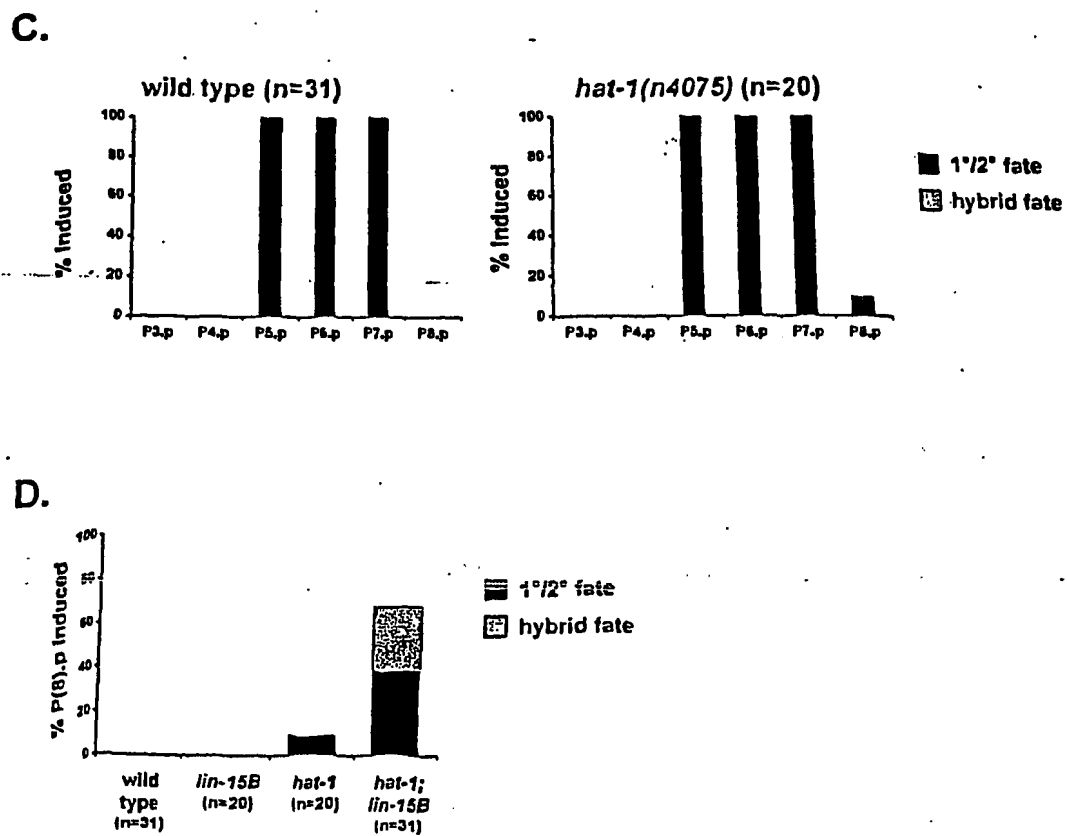


FIGURE 11B



hat-7 genomic sequence

TTGTTTTCGGATTTTTGTGTGCTTCGTAGTTGCTCCGATGATGCCGGATTG
AACATTTGAATGTAACATTTGAATTTTGAATTTGAAGGAATTCATTTGAATCTA
AAGCTTGCAGGGTCAAGACCGATACATTCTTGCAACACATGACTCGAAAGTA
TGTAGGAAAAATTGAAGTTGGAACTTGAATTTGATGAAAAAGTACAGTAA
TCCATTCTCTCTTATTTGCGAACTTTCTTCGATTTTGTATTTTCTAGATTTT
TTAAGCTAAAATTTTGTGTTTTATTTTCATTTTTCATGCTTTTCAATTTGCGTT
TTCAACAAAATTATGTTTTTCAGAGAAAATCTCGTGAACAATAACTCGGCTAC
TGTACCATTTAAAGGCGCAGACCTTTTCGCGCAGCATTGATTTAAATTTTTT
GTTCTGTTGGCTCAACAGTGCAATGGACATCTAGATATCTGAAATTTTACCACT
GAATTCAGTTTCAATTTTTAAGCATCTTCAAAAATTTGCGTTTTCTTAATTTTCT
TGTGATCGTTTTTTTTTGAAGTACAATCGTACATTATAAATAACTATTTTTCT
AATTCGAATAATTTAATTCAGATCATTTCGCAAAATAATTGCCTTGAAACGT
TATGCCGCGGTCAATTTTCAACGACCCTTGTTATTCTTTTTTGAATTGCCGCC
CTTTTTCCCTGTGGCCGGCGCAGTGCGGCCGAGGTTGGTTTCTAGGCCAG
CCGGCGCGTTTATTTTTTTCGAGCATGATTTTCACAATTATTTCTTGCATTTTT
AAAGTTTTTATTGATAAAATAGTAAACTAACAAACGGATAATATTATTTAAA
ATTAAAAAACTAGTTTGTTTCAATTTTGGATCGATTTTAGATGTTGTTTCATGGA
TTATGCACGCAAGAAAGTACTATCGTTTACATTTGATTGCTATATTATTGAAT
ATTGAATTTTTCACACAAAATTGTACTATTTCCAGATATTTATCATGACCGAG
CCGAAGAAGGAGATTATAGAGGACGAAAATCATGGAATATCCAAGAAAATAC
CAACAGATCCCAGGCAATACGAGAAAGTTACAGAGGGATGCCGGTTATTGG
TCATGATGGCTTCAACAAGAAGAAGAAAGTTAGTTTTTACATCTATTTAAACAC
ATTTTCCAATTATTTTTCAGGATGGGCCGAAGTTATTTCAAGATGCCGAGCTG
CAAATGGTTCAATTAAATTCTATGTCCATTATATCGATTGCAACCGAAGACTT
GACGAATGGGTTTCACTCTGATAGGCTCAATTTAGCGTCGTGTGAGCTACCA
AAAAAAGGAGGAAAGAAAGGAGCACACTTGCGGGAAGAAAAGTGAGAAATC
TATAAACTTTTCAAAGATTTTAAATAGTTTTATCAATTCATAATTATTTCACTC
GAGATTCGAATGAAAATGAAGGAAAGAAAAGCGGCCGAAAACGAAAGATTC
CACTACTTCCGATGGATGATCTCAAGGCGGAATCCGTAGATCCATTACAAG
CAATTTCAACGATGACCAGCGGATCTACTCCAAGTCTTCGAGGTTCCATGTC
GATGGTCCGCCATAGTGAAGATGCAATGACAAGGATCCGAAATGTCGAATG
CATTGAAGTAGGAAGATCACGAATTCAGCCATGGTACTTTGCACCTTATCCA
CAACAATTGACAAGTTTGGATTGTATTTATATTTGCGAATTTTGTCTGAAATA
TCTAAAGTCGAAAACCTTGTCTGAAACGGCACATGGTGAGTGTTTCGAGTTAT
AGAAAATGACCGAATATAAATAACTGTTTTCAAATTCAAAATTTTCAATTTT
CCAAAATGAAAGAATCGGTGAATTCGAAAAAATTCGAGTTCTTGTGTGTTTT
TGGCTGAATTTTTCGGTTTTTCTTGCTTTTTCCGTTGATATTAGTTTTGAAACA
ATGTTTTTAAATTTTTCCGGCATCGAAAAAATCGCAAATTCTGGGAATTTGC
TCCAAAATTTGCATTTTTGAAATACTTTTTTTCGAAAACGAAAAAAAATTC
CAAACGGTGTTTCAAACCAAATTTATCGTAATCAAAAAAGTTTCGCAAATAGG
CCATTATTCTGCGTGGGAATTCAAATTAATAATCAGCTACTTTTTCTATTTTGC
AAAATGGAAAAAAACGTAAAAAATAGACAAATTTTAAATTTTTTAAACAATTA
CATTCGGTCCATACTCTTCATTTTCTATCATTTAATTAAATGCCCAATTCTAA
TTAATTTTATTTTCAAGGAAAAATGTGCAATGTGTACCCACCTGGCAATCAAT
CTACAGTCACGATAAACTTTTCAATTTTTTGAATTCGACGGCCGCAAAAACAAA

FIGURE 12.

AGCTATGCTCAGAATCTATGCCTGCTTGCCAACTTTTTCTGGATCACAAGA
CTCTTTACTATGACACGGATCCATTTTGTCTATGTGCTAACCGAAGAAGA
CGAGAAGGGTCATCATATAGTTGGATACTTTTCAAAAGAAAAAGAATCAGCT
GAAGAATATAATGTTGCGTGTATTCTTGTGTTACCTCCATTTCAAAAGAAAGG
ATACGGAAGTTTGCTCATCGAATTCAGCTATGAACTCTCGAAAATTGAACAG
AAGACAGGATCACCCGAAAAACCACTATCAGATTTGGGACTTCTCTCATATC
GATCGTACTGGTCAATGGCCATCATGAAAGAGCTTTTCGCATTCAAAAGACG
ACATCCAGGCGAAGATATCACAGTTCAGGACATTTCACAAAGTACATCGATT
AAACGAGAAGATGTTGTGTCAACGTTACAGCAACTTGATCTATACAAATACT
ATAAGGGATCATACATAATTGTGATTAGTGATGAAAAGCGTCAAGTTTATGA
GAAACGGATTGAGGCTGCGAAAAAGAAGACACGAATTAATCCAGCAGCTCT
GCAATGGCGACCCAAAGAGTACGGAAAGAAAAGAGTGAGTTTTTTTTCAATCA
AAAATTCGTGTTTACGGCTAAAAACTGAAAATTAAATTAATTAATTCGTG
ATAACATTTTTTTTTCAAAAAACCAAAAAAAACAATTTCTGTTTTTGGCAGAAC
CAAAAAAAAATTTAAAAAAAACGGTTTACGCCCTATTTCATACAAACAACA
GAAATTGCATTTTTTTGAGCAAATTTGACCCTACAATTTTTTCCAGTTTTTG
CTCTTTTTCAAAAAAAACACCTAAACACTGGAAATACTAAATACTAAGGAAA
AAAATGGAAATACTGGTTTACAGTGTCAAAAAATTGAAATTTTCTAATAAAAT
CATTTTTCTTTTTACTAAATTTATCAAAAAATTTATAACTCAAATCTTTCAGTTTT
TGCGAATTTTTTTTCGAAAAAACGAAAAAAATAAACCTAATTTTAACCAAATT
GTAATTTTGAAAAATCTGGAACGTCCGGAAAACTGAAAAATTAAAAAAAAG
TTTTCAGAAATTTATTTTTAAAAAACCGTTTTTTTAAATCAAATTTTGTATATGT
TGATGAGAAAAAAAATAGAAATCAATGTTTTTAAGTTTTAAAAGAAAAATTTA
TTTTAATTATTTTAGTTTTAATAAGGTATTTAAACAGTAACAAGGATGTCGGTT
TTTCGATTTTCCGAAAACTAAAAAATTGTCTTTTTCGATTTTTTAATCGAAAA
AAAATAGAAATATTTTCACAAAACATACTATTCTTCTAAAAAAAAGAATAGTG
GGAGATTTTAAATAATTTTTGAACTCTCGCAATTTTTTTCGAAATATCGAAAA
TCGAAAAACCGGCACAAAAGCAAAAAGTCTCCGGGAATATATCTTTAAATTA
TTTTATGAACTTTTTTTTTCAGGCGCAGATCATGTTCTAGCAACAACGACATGT
GTTCTCGCCACGACGATCTCAACCTGTACATTAATAATAACACTCCGTTTTTA
TCTCGCATCTACACACCGAAAAAGCTTACGCTATCCCTTTATCATTCCCACAC
CGCTCAGAGAGCGTACGCCTCATTTCATTTCATTTGTTCTGTGTAATAATTTG
ACTTATTAGTCACTTATTTTTTTAATGAAATTATTCTTGAATTTATAATCTTCT
TGTTGCAGTTCAAATAATTAATAATTATCATATAGACAAGTAAGTTTATAACT
GCAAAAGTGAAGTTTTCTAATCATTAAAGCGTTCTGAAGATATTCGGCAACCG
CCTGAGCGATCAGATCACGGCGGGAAACGAGTTGAGGCGTAGACATGCTTG
CAGCCAGTGACAACCTGAAAGATATTCAAAAAATTAATTTCAGGACTCGAAT
TTTTAACAATCTGAATAAAAAAATCCAAAATTGTATATTATAGAGTTTTTTGAA
ATCTAAGCGAAAGCGCGCTCCAATGTAAAACGAAAAGTGCTCCGCCCTAA
ACGTTGGGTCCCGTTAGGAATTTGTTATTTTTTTCGGTTATTTCTGACTATATT
ATAATTTCGAAACGACAAGTATTTTAAACATCATTTTCGACATAAAAAATATGT
AAAACAACAAAAACAATCGAAAAAATAGTGAAAAAGTTTGAATTTACAGTCT
CGCCGCCTCCTACCGAGACCTAACGTTAGGAGGCGGAGCGTTTTCTTTGG
CATTGAAGCGCGCTTGCTGCGGCCCCATAATTAATAACTTACAGCCTTTGCA
AAGTCCTTCTTCTGTTTCATCCTCAATCTCGTCAATGTATTGATTGGACAACCT
CTCAATCTCGGACTGTTCCGCATTTTCATCCTTCAATTTTTTGATTGAGCCT

FIGURE 12

TGAATTGAGCCACCTTCTCCTCTCCGAAAGCCTTAACCGAATACTCCTTACA
AGCTTCTTTCAACTTGCCCTCGGCCTTCTCCTTGGCATCTC

FIGURE 13

hat-1 ORF

ATGACCGAGCCGAAGAAGGAGATTATAGAGGACGAAAATCATGGAATATCC
AAGAAAATACCAACAGATCCCAGGCAATACGAGAAAAGTTACAGAGGGATGC
CGGTTATTGGTCATGATGGCTTCACAAGAAGAAGAAAGATGGGCCGAAGTT
ATTTCAAGATGCCGAGCTGCAAATGGTTCATTAAATTCTATGTCCATTATAT
CGATTGCAACCGAAGACTTGACGAATGGGTTCAGTCTGATAGGCTCAATTTA
GCGTCGTGTGAGCTACCAAAAAAAGGAGGAAAGAAAGGAGCACACTTGCG
GGAAGAAAATCGAGATTGCAATGAAAATGAAGGAAAGAAAAGCGGCCGAAA
ACGAAAGATTCCACTACTTCCGATGGATGATCTCAAGGCGGAATCCGTAGA
TCCATTACAAGCAATTTCAACGATGACCAGCGGATCTACTCCAAGTCTTCGA
GGTTCCATGTCGATGGTCGGCCATAGTGAAGATGCAATGACAAGGATCCGA
AATGTGCAATGCATTGAACTAGGAAGATCACGAATTCAGCCATGGTACTTTG
CACCTTATCCACAACAATTGACAAGTTTGGATTGTATTTATATTTGCGAATTT
TGTCTGAAATATCTAAAGTCGAAAACCTTGCTCTGAAACGGCACATGGAAAAAT
GTGCAATGTGTCACCCACCTGGCAATCAAATCTACAGTCACGATAAACTTTC
ATTTTTTGAAATCGACGGCCGCAAAAACAAAAGCTATGCTCAGAATCTATGC
CTGCTTGCCAACTTTTTCTGGATCACAAGACTCTTTACTATGACACGGATC
CATTTTTGTTCTATGTGCTAACCGAAGAAGACGAGAAGGGTCATCATATAGT
TGGATACTTTTCAAAGAAAAAGAATCAGCTGAAGAATATAATGTTGCGTGT
ATTCTTGTGTTACCTCCATTTCAAAGAAAGGATACGGAAGTTTGCTCATCG
AATTCAGCTATGAACTCTCGAAAATTGAACAGAAGACAGGATCACCCGAAAA
ACCACTATCAGATTTGGGACTTCTCTCATATCGATCGTACTGGTCAATGGCC
ATCATGAAAGAGCTTTTCGCATTCAAAGACGACATCCAGGCGAAGATATCA
CAGTTCAGGACATTTACAAAAGTACATCGATTAAACGAGAAGATGTTGTGTC
AACGTTACAGCAACTTGATCTATACAAATACTATAAGGGATCATACATAATTG
TGATTAGTGATGAAAAGCGTCAAGTTTATGAGAAACGGATTGAGGCTGCGA
AAAAGAAGACACGAATTAATCCAGCAGCTCTGCAATGGCGACCCAAAGAGT
ACGGAAGAAAGAGCGCAGATCATGTTCTAG

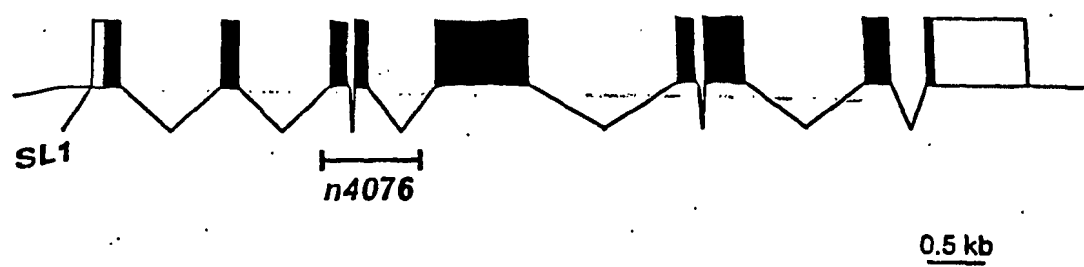
FIGURE 14

HAT-1 protein

MTEPKKEIIEDENHGISKIPTDPRQYEKVTEGCRLVMMASQEEERWAEVISR
CRAANGSIKFYVHYIDCNRRLEWVQSDRLNLASCELPKKGKKGAHLREENR
DSNENEGKKSGRKRKIPLPMDDLKAESVDPLQAISTMTSGSTPSLRGSMVM
GHSE DAMTRIRNVECIELGRSRIQPWYFAPYPQQLTSLDCIYICEFCLKYLKSKT
CLKRHMKEKCAMCHPPGNQIYSHDKLSFFEIDGRKNKSYAQNLCLLAKLFLDHKT
LYYDTPFLFYVLTEEDEKGGHHVGYFSKEKESAEYENVACILVLPFFQKKGYGS
LLIEFSYELSKIEQKTGSPEKPLSDLGLLSYRSYWSMAIMKELFAFKRRHPGEDI
TVQDISQSTSIKREDVVSTLQQLDLYKYYKGSYIIVISDEKRQVYEKRIEAAKKKT
RINPAALQWRPKEYGKKRAQIMF

FIGURE 15

A.

epc-1

B.

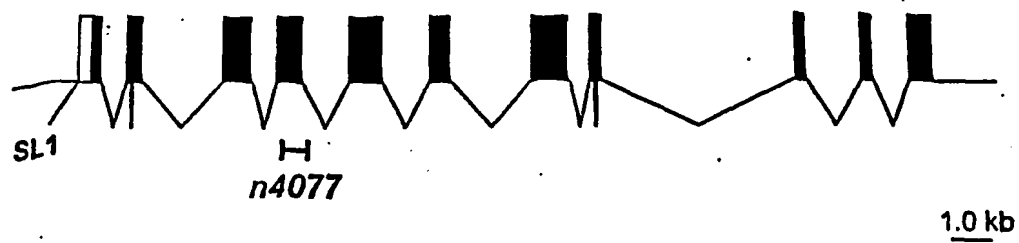
ssl-1

FIGURE 16

epc-1 genomic sequence

TTTCAAAAAAAAAAATTACCTCGTCAATTTCACTCTCCTCGATGCGATGATT
ATCCTCGTCCATTTTACCTGAAAAGTGTGATTTTTTACGAATAAAATTATTTT
CAGATACTTCTAGAAAAAAAAAACTGAACGGAATGTTACGAAATTAATTTTCA
AAGTTGCGAACTGAATTTTCGACAAAAAGTTTCACTGATATTCATTTCAAGC
ATATTGCAACGTTTTTAAATTAATTTCTAAGAGAAAAAACTGCAAAACAATTC
GAAAATAATTTTTACAAGTTACTTTTCGAAAAAGTAACAAAAATCCACTAATG
AACAGAAATTTTTGAACAAAAAGAGCTTCTCAGGCTATTTTTGGACGAATAT
TTTAATAAACTTTAAAAAATCAACGAAAATCCCCTAAAAATCGCTGAAAAT
TCCAAAAATTAAAGTTCATTCTCGACCACACCTCTCGTAAATCAGCACGAGA
CTCACGCAACGCGACCGCGCCGCACTCAACGGCATTGAGTAATGCGGAGC
GGCAGCGTCGCGTCGTCTATTTGTGTGTGTGTGCGATTGTGTGTGGTGCGA
CGTGGCCGCTCTGTGTGCCTCTCTAGTGAGTGTTTTCCGACGAGAGACAAC
ACATTTTCGAGAGACGAAGAGAGTGGCGACGAGGAAGATAGTGTGGTAAGA
GGAGAGTGTGCGCGAGGGAAAGAGAGCAAAGTGTGAGTGTCTGTGAGAAG
AGAAGGAGACCCCCCCCCCCCCGCGCTCAACCAGTCGATAGTTGGCCTGA
GTGTAGGGCCTTCTGTTGTATTCCACTGCTAACCCCCCCCAACACACAAAA
AGACTCAAAAAGTACTGCTTAAACACAGTGCTCAGCTCATTTTCATTTTTGAT
TTTTATGCTCGCCGTCATCGGCGGATGAATTCATCGCAAAGTCCGTGGCGA
TTCAACACGTGCGGCGTCTCGCCGCTCTTCTTAACCGTAGTTACAACGTG
GGAGTACAGAAAGATCGCCACTACTTCGAAGGCGTTTCGAGCCCGGGCGC
TCGACTCGAACCGGTCTATGACTGTATACTGGGGCCACGAACCTCCGGACC
TATCAGAATGCAGTGTTGGAAACCGGGCGGTGACACAAATGCCGTCTGGCA
TGGAAAAAGAAGAAGAACAGGTTGGTTTTTGGTGGATTATGGATTACTGCTC
CATTTTGAAATTTTTCGAGTTTTAATGTCTTTTTTCGAATTCCTGGTGCTTTTT
TCTATCCGAATCATGTTTTAATTCGTTTTCCGACTACTTTGAAGAATTTTCA
AATTTTTGATCCCTGATGACGTCATAATTTTTGTCTTTGCCTTTCTGGATCGC
TTTTATAGTTATTTTCATTTTTTATTTCTTTTTTACACTTTTAACTTAACAATTC
TCTTAATTCATCCTATTCTATTTAATTTAAGTTTTGATTTTTGATTTTGATT
TTCTCTTTTCTTTTTAGCCGCCGGTGGGCCCTTATTACAACCTCTTAATCAT
AAAAAAAATCAGTTTAAGCAGTTATACATAACTCTTATTATGAAAAATCGTTA
TTTTTCGACGGAACTTCATACTTTGAATTTATTTCCAATTTAGATTTTATTTT
CTCAAAGTCAGCTCAATTAACATACTTAAATGTTTTGTCTACCCGCAAAAT
GTTTTTTTTTAATATTTTAATTTCTATTTTAATTTTTGGCTTTAAAAAATCATTTT
GCTAAGCCTGAGATGAAGGCGAAATCTCGAGAAAAAGCATTTAAAAAGTAAT
AAATTCGGTTAAAAACGACTTTTTCTATCACAGAAAGTGTTCTCTGAGTGCTA
ACAACCTTCTTCTGTCCAAATTTTGACACAATTTCCCAATTATGCCGACTTAT
TACACCTTTTTCCGTCAATCTTCTAGTTTTTCCCACCCTCTTGACCCCTGGTG
ACGTCATTTGTTTGTTCTTCTTCCAAGACATGCCCTGTGGGGTATTTTTTCTC
AAAATTTTTGCAAATTTATTGGATTCTAAATAAAATTCAGGAGTCTAGCACC
AGGAATAATAATGCAAATTTGAAAAAAAATTAACAGAAATAATGATTTTAA
ATGATTATTTAAATTTTAAATTTTAAATTTCCAGGAAAAACACCTGCAAGAAG
CGATTGGTGGCGAGGAAGCGAGTAGATCGGGTATTCAGGTGAAGCATGTCA
TTCCAACCTCAAAAAGTCGACCGAGTCGAAGATCAACGCTATCACTCCACTTA
TCACAACAAGAATAAAATGCACCGTTCAAAGTATATCAAAGTTCATGGTGAG
TTTTTTAACC AAAATTTTCGGCGAAAATAATTTAATTTCCGGTTTTTTGAAATT

34/92
FIGURE 16

AATTTCCGCTTGGGTTTTCTTGATTTATTATTTTTGAAATTCCTCTCTGAAT
TCGAAAGAAAATAACTTGATTTTTAGACTTCTGGCTAAAACCTTCAAAAAT
GTTTGTTGATTGGTTCCAAATTTTCGCCTGATTCGGAATTCGATGTGACAAA
TTCAAAAAAAATTCCTGATTTTATATTCAAGCTTTGTGTTTGTGTGTTCTTT
TTGGAGCGCGCTTGCATCGTTTGATTTTCTTCGTCTTTTTTAAATTTATTTTC
GCTTGTTTCATTCATTTTTGTGAGTTTTTTTTCTGCCAAAATGAATGAACTG
GTTTAAAAAATTGAATTCGGCGAAAATAAATTTGAAAAACGAAACAAATCAA
ACGATGCAAGCGCGCTCCAATGCGATTTTTTTGGGCGCGGAAATTCGTGAT
TTCAAGCTTAAATATAAAATCAGGTATATTTTTTCGACTTTTTTCACGTTGAAA
TTCGGAATCAGAGGAAAATTTTGAGTCAATCAAAAATATTTCCCAGATTTTCG
GTATCTTTAATGCATCAAAAATGAACTTTCACCCCCATACTCCCAGAAAAATA
AGAAAACAAATTGCGAATATTGTTCCCTGATCAAATTTTTCTTTTTTAACT
ACACTTCTCTGTTTTGAAGTGAGAAAGTACATTTTTCTGCGTTTCTTATCAGT
TATCATTTGAAAAGGATCAGAATTTGATGACGATATATTTGTTTAGTTACCTC
CCTTTTTCTGAACAGTTTTTGCGAAAAAAGGAGAAAAACCGGAATTTCTAT
GAAAATGTGATTTATTTTCAGCCTGGCAAGCACTCGAACGAGACGAACCCG
AGTATGACTACGACACAGAAGATGAAGCATGGCTATCAGATCACACTCACAT
TGACCCGCGCGTTTTGGAAAAGATATTCGACACAGTGGAGAGCCATTTCATC
GGAGACACAGATCGCGAGCGAAGATTGGGTGATTAATTTGCATAAATGTAA
GTTGACGAAATTTCCATTGAAACCCCCCCCCCAAAAATATCGTTTAATTG
CAGCACTGGACTCATCAATCGTGTACGAAATATACGAATATTGGCTGTGAA
GCGAACATCGGCTGCGACGACGTCTGGTTGTGTTGGAGTCGGTGGATTAAT
TCCGAGAGTCAGGACAGAATGTGGAAGGTAAGAATTTGACTATTTTGAAC
GAATTTCTGTGATGAACTTCTCTAAAACCTTTTAAAGTTTTTTATGGCGGTTCA
AAATTTCGGAAAATTTACACTGATTTTAGCTAAAACCTTGAATTTTGGTCATTT
GTCCGTGTACATCTGTCCGAAATCGACTTTTTTTGGAATTATCATCCTTTAT
TGCACATTTGGCTAGTTTATCTCATTTAATTTGTTGATTACTAAGGTACATTT
AAAGCCAATAGGTAACCAACCAAAAACCTATCATAATTTTTCTACACTTTTTAA
TTTTCCGACACTACTTGAATAACCCCATAGTGACCAATTTTGATAGTTTTTG
GCTGGTTACCGGCTTTAAATGTACCTTATTAATCAACAAAATTAATGAGATA
AACTAGCCAAATGTGCAATAAAGGATGATAATTCCATAAAAAGTCGATTTTG
GACAGATGTGACACGGGCAAATGACCAAAATTCAGTTTTTAGCTAAAATCA
GTGTATTTGTTTCGAAGTTTTGAACCGCTATAAAAAAATTTTTGGAATGCTTT
TGGCAAGTTTCATTACGAAATTCATCTATTTCTATACGCAAAAATTAGAATT
TTCAATTA AAAATTCATTTTACAGGATGGACAAGGTGTTATCAATCCGTACGT
TGCATTCGTCGACGTGCCGAGAAAATGCAGACTCGAAAGAATCGGAAAAA
CGATGAAGATTCGTATGAGAAGATTCTCAAGTTGGTACATGACATGTCGAAA
GCTCAACAGCTCTTCGATATGACTGCCCCGACGAGAAAAGCAGAAGCTCGCG
TTGATTGATATGGAATCGGAGATTTTAGCGAAACGAATGGAGATGTCAGATT
TTGGTGGTTCTCCGAGTTCGTTCAATGAGATCACCGAAAAGATTCGAGCAG
CAGCAACGTTGGAAGTCGTGAAACCACCACTGGCAGAAATCAACGGATCAG
ATGAAGTGAAGAAGAGGAAGAAGCCGAGACGAAAGATTGCTGATAAGGATT
TAATATCGAAAGCCTGGCTTAAAAAGAATGCAGAAAGTTGGAATCGGCCCGC
CGTCGCTCTTTGGACAACACAGTGGAAATGTTCCGACGGTTACAACGAAGC
CAGTTCGAGAGTCGTTGGCGAATGGGCGATTTGCGTTCAAGCGGAGGAGA
GGATGTGTTTATCGCGCGGCTCTCACC GTTACAATGTGCCTACAGCGCCT

35/92
FIGURE 16

GCTACAGTACCTCCAGTACAGACTCAAGCAGCAGTGGCTTCATCATCATCG
TCAAAATCAACGGATATGGTGCCGTGGAACATGAAGTTCTTTGAACTTTTG
TTCGGGATTACAGGATTCAGTTTCTCGATCTCTTGGCTTTGTACGCCGACG
AATGGGACGAGGTGGGCGAGTTGTATTGATCGGATGCCTCGCAATCGAG
ACGACAACGACGAACGCACTTCGACAGATCCATGGGCCGAGTATTGTGTG
CGGATAGTTCAAGGTGAGATTTTTGAATAAGAATCTTAATTTACAGAGATTT
GGTTTTTTTCGCTGCTTTTTCTGTAATTTGTGGTATTTTTCTCGTATTTTCA
ATTA AAAAACGGGTTTTAAATAATTTTAACTGAAATTTGCTAAAAACCAAG
AAATTTTATTAAAAAATGCAACAAAAAAGACTGGAGGCACCACCGAATG
GAGAACAGGAGAACCCAAAACACGCCCATTTTTCCGTGCCGGGCGGCGA
AAATTTTTGCAGAATTTGCTGCAATTTTTCGTTTTACAAACGAAACAACGAAG
CTCTGAATGTGTTATTTGGAGCTTCGTTGTTTCGTTGTAAACGAAAAATT
GCAGCAATTTCTGCAAAAATTTGCGCGCGGCACGGAAAAATGGGCGTAGTT
TTAGGTTCTCCTGTTCTCCTTTCCGGTGGTGCCTCGAGTCTTTTTCGCATTCT
AATGAAATTTCTTTGTTTTTAGCGAAATTTAGGTTAAAATTATTTAAACCC
GTTTTTTTTTCAATTGGAAATGCGAGGAAAAACCACAAAATCACAGAGAAAG
CTTTTGGATTTTTTCGCAGCTTTTTCTGTGATTTTGTGGTTTTCTCGCATTT
TCAATTGAAAAAAAACGGGTTTTAAATAATTTTCACTGAAATTTGCTAAA
AACGAGGAAATTTATTACAAATGCAAAAAGACTGGAGGCACCACCGAAA
CCGAATGCAGCTCAGAACAGGATTTACCAAAACAGGATGCAGTAGGCGGAG
CCAATTCGCAACCACCGCATGCTTATTTTCGCATGCCTCGCACGTTTTTTTT
CTCTTGAAACAATGCAACAATATCAAGGAAAAAACGTGCGAGACTTGCGAAA
TAAGCATGCGGTGGTTGCGAATTGGCTCCGCCCACTGCATTCTGTTTTGT
AAATCTGTTCTGAGCTGCATTCTGTTTTGTTGGGGCTTCCAGTCTTTTTGT
GCATTTTAAATGGAATTTCTTCGTTTTTAGCGAAATTTAGGTTAAAATTATTT
AAAACCCGTTTTTTTTTCAATTGGAAATGCGAGGAAAAACCACAAAATCACA
GAGATAGCGAGGCCCCACGAAAAGGGGAGCAGAACAAAAAGGGGGGGG
GGGGGCTGGCACTGTGCCAAACGCACAAAACGCTTTTTATTCTTATTCAACG
CACGACTTTGTTATAACCACACTCCGTTATTACGCATCGCGCGCTGTTAGC
GTGAAAATACAAAAAACGTGCTGCGTTGAATGAGAATAAAAAAGCGTTTTG
TGCGTTTGGCACAGTGCCAGCTCTCCTTTTCGCAGATCCCCTTTTCGTGGG
GCCTCAGAGAAAGCTGCCATAAACTTTTTCTTCGCGCTAAGACCAATACCA
ATAAATCCTTGCGCCTTTAATATGCAAACTATATTTTTCTTCCAGAACGTTCC
GTGCTCGAAACAGTTGCTTGGTACCGAAGAAGAAACCGATGATCTAAGCC
CGAAATCTCTGTATTTGCTCGCAGTAATCGGTTTCGCATTCAACGATGATGA
AACTGAACGGGAATGGACTTCAAGATGCCAACAAATCATCGTGGAGAGATAC
AGAGGTGGATGATGAGCTGAAAAAGCGGGAAACAACGTCTGAAAGTGAGAT
TTTGAACGATTTACCTGGGAAAATAGATTATTTTGGGCCTATTTAATTATTTA
ATTGCAGAATTTACCGAAACCACGACGAATGGAAGTACCAAAACACACACA
GAATCGGATGATAGTGAAGTTGAACGGATGGAGGTTGATGATCAAGTTGAT
GAAGCTCAAATAACTGTATCATCATCAAAGACGATGGAATGAATGGAATG
ATAAGAACGAGGATGAAGAAGATGATGATGATGATATGGATGTAGATGAACA
TCAGAGTGTGCTGGGTGTGGATCAGCAGGAGGAGGAGGAGCATCACCAGC
AAAAAGTTCCGGCATCAAATGAATGGTGGTGGTGGTGGTGGTGGAGTGGTAA
AACTGAAACCGCCGCTGCAAGAACTTTCCGCCCGCTTTCCGGGAAACGGAA
GAGCGGACAGAGCGGAACCGACGCCGTTCCGGCAAAGGTAGTGAGGCTT

FIGURE 16

TTTTTTTAAATACTCGAAAAAGAAGGAAAAAATCCCACTTTTAAAAATACGAT
TCTTAAAAATGCGAATCCCTCCAAAATGAGAACTCTGATTGGCCAGGGAGC
TCTCATTTTGAATGGAAATTAGTCAAAATTGAAAAATCCCGTTTTTTTTTAAAG
TTGGATTTTTCATTTTCTCGCGATTTTTTCCGCGTTTCTGTGTCATTCTGAA
TTTAAACATTTAATAAAATTAATAATGTCTGGAATATTGACAAATTATGCTTCAAA
TTTTTTGCGCGGGAGTTCAAAAATAATTTGGCCCTTTTTATTTTTATTTTGCA
AAAATATATAAAAAATCATTTTAAAAAATTTAGAAACATTTTTTAATTTTTTAA
CAGTTATATTGCTATATTGGGACGGTATTCTGTCATTAAACTTGGTGTGTC
GAATTTTTTTTTATTGCTTTATAAGACTCAAAATTGTCTGAAAACACCGAATTT
ATAATGAACTTCTTGGAACCTCTCAAAAAAAGTTATGACGGCTCAAAAAA
TGACCTAAAATTTGTTAAAATTTGAAATTTGACTTGTGCAACGGCTGGAAAC
AATTTTTTTTTTTGAAATCACCGTCAAATTTTGAATATAAAATTTAATTATTTTG
CGTTTTCAACTCGATTTTTGGTATTTTCAAGTCGATGGACGGCAAGATTTGG
TTAAAAAATTAAGCCGTCCATTTTCTCGCCGTCCATTGACTTTAAACTACC
TAAATCGAGTTGAAAACGCAAGATAATTGACATTTATACCCAAAATTTGACTG
TGTTTTTAAAAAAGTTAGTTTCCAGCCGCTGCGACAAGTCAAATTTCCAATTT
TAACTATTTTAGGCCATTTTTGAGCCATCATAACTTTTTTTGAGAAGTTTTT
AAGAAGTTTCATCATGAAATTCGGTGTTCAGACAATTTTGAATCTAATAAA
GTAATTTTAAAAAATTCGACAGACACCACCTTTATAGCAATTTTGAATTTTTT
TTAAACTTGTCTTGAAAAATCTTGAAAAAAGTCGAATAAATTTCCATTTTCT
ATTTTCTTTTTGCGAGATGTGCGGAACGGTGTGCGACTCAGATGATTGGAGA
GAGCCGAGTGGATCACCATCAGAATCGAATTCATCAACCGAATGGGGTGGC
TATACGCCACAAGAACAGCATGCAGTTGTTGTTGCCAACGCGGTAGCTGTC
GCTTTCAAGGAAAAATTGATGAATGGCGTGGATGATGATGATGATCAACAAC
CATCGCCGGCTAGAGGAGCACGAGATCATTCCATCAAAGAGTTCGTTAGTT
TTTCTTTGCTTTTTTTTTTTTTGATTTTTGAGAGCAAATTTGAAAAGTTTTACA
CGGTTTTTAAAAAATGTTGAAATTAATAATTTGTTGAGAATTTGATTTGAGC
AAGTTTTATTTTTAAAAAATTGAATTTTTGAGAAAATTCTGAGTTTTCTTTTTAA
AAAATTGAAATTTTCAAGAAATCTGAGTAGCAAGAATCTTTAAGATCCTTAA
TTTCTATGCAAGAATACGTAGGAGTTTTACTTTGCTCAGGAAATTTATTTTTT
GTCAGAGGAGTATATCCGAAAAAGAACAATAAATGCACATTTCTCAAAAC
GCGTATTTTTTTTTTCAAGTTGATGTCAACGGTAACACTGCTGGAACGGAAAA
AGTTCATGATGCCGTGCAATCGGTCTATTAATTTGAACTCTCTGCTGCTGC
TTCTGCTACTGCTGCTACTGCTGCTCATCGCCAATTTCAATCCTCCTGAGA
TTTTTGTGATGGTCATTCATTGTTTTGTGCATATCTCTCTCTCTCTCTCTC
CCATGATTTCTCAAATATTTCAATGTATTTACACCCCCACTCTGTCCGCTGCCT
AATCCCCGACCGAATAATCAGATTGCTGGAAAAATCTGCGATTCTTTAATA
TTGCAACCAACCAACCAATAATATGTGTCTCATCATCTCGGTACTCTCACTT
GAGCCGTGTTTTCTGTAGTATTTTATTCTCTAAAAAATAATCATTTTTAATATA
ATATACGTACACATTTATATCTGTAATATATATTTTTAAAAATGATCCCCCT
CCCTCCATTGTTGTTTTTTTTTCTGTGGGTTTCAAGCTTTTGAAGTGTGAAA
AATCTCATCCCATCATCATTTTCTATTGTTTTTTTTTCAAGTTGAAATATCTA
TTTTATCTTTTTCTTTTTTTTTTCAATTTTTTTTTTTGATGGTGGGGGATTCATT
TTTGGTCCCGCGAAACGCCCGCCGCCGCAATCCCACTCTCTCTCTCAGT
CTCTTCTTAATGATCTTCGAAACTATTTTTATTTCCCTCATTAACAATTACGAG
GTCGTCTTTTTTTTTTCCCCACCCCCCACTGTTTGGTGTAAATTTTGTGTTGCG

FIGURE 16

GGAGGTTTTTGTGTGTGGATTTTGGATTTTGGATTTTCAACAAAAA
TTCCCCCGAAATCAAATTTTTCCCATTTCCCCTCAATATTAGTACTGTTG
TATAAATAAACTTGCTCTCTCTCTCTCTCGAAATCTCCTACTATTATTTTT
TAAAAGATTTTCCAACAAAAATTCAAAAAACCACACAAACGACCTCTCTGCA
CGCGGTAATCCTCTCTCTTTTGTCCCCCATTTTCTCTGTTTCTCTTTTTTCT
ATCCCCTATACCTGTGATTGGAATATC

FIGURE 17

epc-1 ORF

ATGCCACTACTTCGAAGGCGTTTCGAGCCCGGGCGCTCGACTCGAACCG
GTCTATGACTGTATACTGGGGCCACGAACCTCCGGACCTATCAGAATGCAG
TGTTGGAAACCGGGCGGTGACACAAATGCCGTCTGGCATGGAAAAAGAAGA
AGAACAGGAAAAACACCTGCAAGAAGCGATTGCTGCCCAGCAAGCCAGTAC
ATCGGGTATTGAGCTGAACCATGTCAATCCAACCTCCAAAAGTCGACCGAGTC
GAAGATCAACGCTATCACTCCACTTATCACAACAAGAATAAAATGCACCGTT
CAAAGTATATCAAAGTTCATGCCTGGCAAGCACTCGAACGAGACGAACCCG
AGTATGACTACGACACAGAAGATGAAGCATGGCTATCAGATCACACTCACAT
TGACCCGCGCGTTTTGGAAAAGATATTGACACAGTGGAGAGCCATTATC
GGAGACACAGATCGCGAGCGAAGATTGCGTGATTAATTTGCATAAATCACT
GGACTCATCAATCGTGTACGAAATATACGAATATTGGCTGTGGAAGCGAACA
TCGGCTGCGACGACGTCTGTTGTGTTGGAGTCGGTGGATTAATTCCGAGA
GTCAGGACAGAATGTCGGAAGGATGGACAAGGTGTTATCAATCCGTACGTT
GCATTCCGTGACGTGCCGAGAAAATGCAGACTCGAAAGAATCGGAAAAAC
GATGAAGATTGATGAGAAGATTCTCAAGTTGGTACATGACATGTCGAAAG
CTCAACAGCTCTTCGATATGACTGCCCGACGAGAAAAGCAGAAGCTCGCGT
TGATTGATATGGAATCGGAGATTTTAGCGAAACGAATGGAGATGTCAGATT
TGGTGGTTCTCCGAGTTCGTTCAATGAGATCACCGAAAAGATTGAGCAGC
AGCAACGTTGGAAGTCGTGAAACCACCACTGGCAGAAATCAACGGATCAGA
TGAAGTGAAGAAGAGGAAGAAGCCGAGACGAAAGATTGCTGATAAGGATT
AATATCGAAAGCCTGGCTTAAAAAGAATGCAGAAAGTTGGAATCGGCCGCC
GTCGCTCTTTGGACAACACAGTGGAAATGTTCCGACGGTTACAACGAAGCC
AGTTCGAGAGTCGTTGGCGAATGGGCGATTTGCGTTCAAGCGGAGGAGAG
GATGTGTTTATCGCGCGGCTCTCACCGTTTACAATGTGCCTACAGCGCCTG
CTACAGTACCTCCAGTACAGACTCAAGCAGCAGTGGCTTCATCATCATCGTC
AAAATCAACGGATATGGTGCCGTGCAACATGAAGTTCTTTGAACTTTTGT
CGGGATTACAGGATTGAGTTTCTCGATCTCTTGGCTTTGTACGCCGACGAA
TGGGACGAGGTGGGCGAGTTGTATTCGATCGGATGCCTCGCAATCGAGAC
GACAACGACGAACGCACTTCGACAGATCCATGGGCGGAGTATTGTGTCGCG
GATAGTTCAAGAACCTTCCGTGCTCGAAACAGTTCGCTTGGTACCGAAGAA
GAAACCGATGATGTAAGCCCGAAATCTCTGTATTTGCTCGCAGTAATCGGT
TCGCATTCAACGATGATGAACTGAACGGGAATGGACTTCAAGATGCCAAC
AATCATCGTGGAGAGATACAGAGGTGGATGATGAGCTGAAAAAGCGGGAAA
CAACGTCTGAAAAATTTACCGAAACCACGACGAATGGAAGTACCAAAACACA
CACAGAATCGGATGATAGTGAAGTTGAACGGATGGAGGTTGATGATCAAGT
TGATGAAGCTCAAATAACTGTATCATCATCAAAGACGATGGAATGAATGGA
AATGATAAGAACGAGGATGAAGAAGATGATGATGATGATATGGATGTAGATG
AACATCAGACTGTCGTGGGTGTGCATCAGCACCGAGCAGCAGCATCACC
AGCAAAAAGTTCGGCATCAAATGAATGGTGGTGGTGGTGGTGGTGGAGTG
GTAAAATGAAACCGCCGCTGCAAGAATTTGCGCCGCCGCTTTCCGGGAAAC
GGAAGAGCGGACAGAGCGGAACCGACGCCGGTTCCGGCAAAGATGTGCG
GAACGGTGTGCGGACTCAGATGATTGGAGAGAGCCGAGTGGATCACCATCA
GAATCGAATTCATCAACCGAATGGGGTGGCTATACGCCACAAGAACAGCAT
GCAGTTGTTGTTGCCAACGCGGTAGCTGTGCTTTCAAGGAAAAATTGATG
AATGGCGTGGATGATGATGATGATCAACAACCATCGCCGGCTAGAGGAGCA

FIGURE 17

CGAGATCATTCCATCAAAGATTGATGTCAACGGTAACACTGCTGGAACGG
AAAAAGTTCATGATGCCGTCGACAATCGGTCTATAA

FIGURE 18

EPC-1 protein

MATT SKAFRARALDSNRSMTVYWGHELPDLSECSVGNRAVTQMPSGMEKEE
EQEKHLQEAIQAQASTSGIQLNHVIPTPKVDRVEDQRYHSTYHNKNKMHRSK
YIKVHAWQALERDEPEYDYDTEDEAWLSDHTHIDPRVLEKIFDTVESHSSETQI
ASEDSVINLHKSLDSSIYIEIYEWLSKRTSAATTSGCVGVGGGLIPRVRTCEKRD
GQGVINPYVAFRRRAEKMQRKRNKNDSDSYEKILKLVHDMMSKAQQLFDMTAR
REKQKLALIDMESEILAKRMEMSDFGGSPSSFNEITEKIRAAATLEVVKPPLAEIN
GSDEVKKRKKPRRKIADKDLISKAWLKNAESWNRPPSLFGQHSGNVPTVTTK
PVRESLANGRFKRRRGCVYRAALTVYNVPTAPATVPPVQTQA AVASSSSSK
STD MVPSNMKFFETFVRDSQDSVSRSLGFVRRRMGRGGRVVFDRMPNRDD
NDERTSTDPWAEYCVADSSRTFRARNSSLGTEEETDDLSPKSLYFARSNRFAF
NDDETEREWTSRCQQSSWRDTEVDDELKKRETTSEKFTETTTNGSTKHTES
DDSEVERMEVDDQVDEAQITVSSSKDDGMNGNDKNEDEEDDDDDMDVDEHQ
TVVG VHQHQQQQHHQKVRHQMNNGGGGGGGVVKLPPLQELSPPLSGNGR
ADRAEPTPVPKMCMTVSDSDDWREPSGSPSESNSSSTEWGGYTPQEQHAVV
VANAVAVAFKEKLMNGVDDDDDDQQPSPARGARDHSIKDSMSTVTLLERKKFM
MPSTIGL

FIGURE 19

ssl-1 Genomic

```

cagctgatgt tgttgatgga aaaatgacgg ctgcaaagaa gccattggct gcaactgagc 60
caaaagtgca taataaataa atgtgtttct aggatcttct aataattttt tttctgtttt 120
ctagctctaa acttgatttt atttcattct tgttctacca aattcccacg gattctacgc 180
tttatgtttc taaattatta ttctttttta tttatatctg cattttcttc taaaaactct 240
ggtcattttc ttgttttttt cttggttaatt ataaaaatta gtcatacaaa tcttgttaaa 300
tatctggcta ttcagtgaac aaaccatttt cgcgtctaaa ttcgaccgca atcaatcgaa 360
aaatggctca aaacgatgcc atctggctgc aacccccctg tcgtctctca attttggtgta 420
ctctctcgca gccacgcacg cgacgcaacg cactcgcgtc gcggtcgcag ttctttttca 480
aatttatcgc gccatttttg ttttgccctca tatttatcgg ctcacgattg attttcgtcg 540
aaaaacgcgc ttaatcgatt cctttttacc tgaaaaatgt tgttccaatt ggaaaaccag 600
ttgaagatcg atgaattttc aagaaaatca ttcaaatagg caaaaccgcg tgaactttga 660
aattcgattt ttgagttttt tgaagaaaat ataattattt catcatttat gttggtcctg 720
ttggtcctca gcatagaaaa ttcggacatg acattagaaa ttcataataa ctgctcccaa 780
tatcgggatt agaacgattt tcagctcaaa atatggaaaa ttggttacat aaaccgcata 840
ttttagcat taatcttgaa cagctatatg gcattaaaaa aaaatatata tatacattgt 900
ttttctctc gaagtttctc tttttgtttc taaaatccgg aatataattt aaaaaaccac 960
ataaatttca atttgcagta cgagttcccc ccgaatcaca atg ccg gca aca ccg 1015
                               Met Pro Ala Thr Pro
                               1           5

gtg cgt gct tca agt act cga ata agc aga cgt aca tca tca aga tca 1063
Val Arg Ala Ser Ser Thr Arg Ile Ser Arg Arg Thr Ser Ser Arg Ser
                10                15                20

gtg gct gat gat cag cca tca act tcg tct gcg gtg gct cca cct cct 1111
Val Ala Asp Asp Gln Pro Ser Thr Ser Ser Ala Val Ala Pro Pro Pro
                25                30                35

tca ccc att gcc ata gaa act gat gaa gat gcg gta gtt gag gag gag 1159
Ser Pro Ile Ala Ile Glu Thr Asp Glu Asp Ala Val Val Glu Glu Glu
                40                45                50

aaa aag aag aaa aag aca tca gat gat ttg gaa att atc act cca aga 1207
Lys Lys Lys Lys Lys Thr Ser Asp Asp Leu Glu Ile Ile Thr Pro Arg
                55                60                65

act cca gtc gat cgg cga att ccc tac att tgc tcg att ctt ttg act 1255
Thr Pro Val Asp Arg Arg Ile Pro Tyr Ile Cys Ser Ile Leu Leu Thr
                70                75                80                85

gaa aat cga tcg att cgc gat aaa tt gtacgatttt ttaaatttaa 1301
Glu Asn Arg Ser Ile Arg Asp Lys Leu
                90

ttactttcct caaatccgaa taattattag atcgcgcttc gcgtttctgc atccgcggta 1361
ttttgccttc ccactgaaaa tagcagattt atcgaatttt tagcttaaaa aaaaaatggt 1421
ttttctgcat ttttcaaaca aaccttttgt aaaacagtga aaatcgaatt tcaaatgact 1481
aaaaatgaatt ttttttttgt ccaactggtg tggaatgggt tgaatttgaa gaaatcagcg 1541
ggatttttcg tattttctga atatttttct attaaaaatc ggtttcaaac cattttttga 1601
cttttgaata gaaaaatatt gagaaaaatc gaaaaatcca gctaacttcc agcttgttca 1661
aattcaaac attccacaac cagtggacga aaaaagttca ttttagtcat ttgaaattcg 1721
atttggttg tttgaaaaat gcaaaaaaaa aatatttttt aaagctaaaa atttgataaa 1781
tctgaaaaaa atctgctatt ttcagtggaa aggcaaaata ccgcgaagcg cagcaagcgc 1841

```

FIGURE 19

gctctaataa	ttattccgct	tcgagaagag	cggtgattat	ttcattgtta	catttcaaaa	1901		
ttatgaatta	atgtttttca	g g gtt	ctg agc agc	ggt cca gtt	cgt caa gaa	1953		
		Val Leu	Ser Ser Gly	Pro Val Arg	Gln Glu			
		95		100				
gat cac	gaa gaa	cag att	gct cga	gct caa	cgg ata	cag cca	gtt gtc	2001
Asp His	Glu Glu	Gln Ile	Ala Arg	Ala Gln	Arg Ile	Gln Pro	Val Val	
105		110		115		120		
gat caa	att caa	cga gtc	gag caa	at gtatgtgaag	ctgaaaaatt			2047
Asp Gln	Ile Gln	Arg Val	Glu Gln	Ile				
		125						
gcaccacaaa	tcaattattc	taattctgtt	ttacag c	ata ctc	aat ggt	tca gtg		2102
				Ile Leu	Asn Gly	Ser Val		
				130		135		
gaa gat	att ctg	aaa gat	cct cga	ttc gca	gta atg	gca gat	ctc aca	2150
Glu Asp	Ile Leu	Lys Asp	Pro Arg	Phe Ala	Val Met	Ala Asp	Leu Thr	
		140		145		150		
aaa gaa	cca cca	cca aca	cct gca	cct cct	cct cca	atc cag	aag aca	2198
Lys Glu	Pro Pro	Pro Thr	Pro Ala	Pro Pro	Pro Pro	Ile Gln	Lys Thr	
		155		160		165		
atg caa	ccg att	gag gtg	aaa att	gag gat	tca gag	ggc tca	aat acg	2246
Met Gln	Pro Ile	Glu Val	Lys Ile	Glu Asp	Ser Glu	Gly Ser	Asn Thr	
	170		175			180		
gct caa	ccg agt	gtt ctg	ccc agt	tgt gga	gga gga	gga gag	acg aat	gtg 2294
Ala Gln	Pro Ser	Val Leu	Pro Ser	Cys Gly	Gly Gly	Gly Glu	Thr Asn	Val
185			190			195		
gaa aga	gcc gcc	aaa aga	gtgagttttg	aagatagatt	ggtgtgtaaa			2342
Glu Arg	Ala Ala	Lys Arg						
200		205						
aaatgaatgt	ttatatattc	actgcaactt	tttcctcacg	agggacgagg	aaaagtgggt			2402
tctaggccat	ggccgaggtg	ccgacaagtt	tcagcggcca	tttatcttgc	tttgttttcc			2462
gcctgttttc	tttcgttttt	catcgatttt	tttcgttttt	tcttaataaa	actgataaat			2522
aaatattttt	tgcatatgct	aaaacaattt	ccaagtaaaa	aaattatgta	ttcagtgggc			2582
aagcagcggt	gaaagtggtc	aatgcaatat	gatggattac	gggaatacaa	aacctaaact			2642
ttttctgaaa	catgatacat	acgctgctta	aatgctgaga	ctacctgatt	ttcataacga			2702
gaccgctgaa	aaagtgttga	ggttttcaaa	attcaaat	tttggtgaaa	aagtcgagat			2762
tttcgcacaa	aaagtgtgaat	tctgaaaaacc	tcaaattttt	ttcagcggtc	tcgttatgaa			2822
aatcaggtaa	tttcagcatc	atatgtatca	tgtttcaaaa	aaagttagg	ttttgtatc			2882
ccgtaatcca	tcatattgca	ttgaccactt	tcaccgctgc	ttgcccactg	aatacatgat			2942
tttttacttg	gaaattgttt	tagcatctgc	aaaaaatatt	tatttatcag	ttttattaag			3002
aaaaaacgaa	aaaaatcggt	gaaaaacgaa	agaaaacagg	cggaaaacaa	agcaagataa			3062
atggccgctg	aaacttgctg	gcccctcggc	catggcctag	aaaccacttt	tcctcgctcc			3122
tcgtgaggaa	aaagtgtcag	tgttattgta	aatctcacaa	gagtcctggca	tgattttctca			3182
aaggcgcatg	gatttattca	gccctaaaat	taaataaatc	catacgactt	taaaggtgga			3242
gttcggaaaa	tgaggatttt	actttaaaaa	gttcaaaacta	gtcccaaattg	ccgaattacc			3302
acaaaaagaaa	aacggaaaaa	aattcatcaa	gtttgaaaaa	aatgcggatg	attttgttga			3362
aattttcaacg	ctcgctaata	ttcctaattt	gaaccgcgct	tttgctcgcg	ccgcactctg			3422
tagaattgca	tccgcgctgt	ttccttcctc	ttccggcgcc	ctacttcttt	tcgattggaa			3482
atgatgaaaa	aatgagacaa	aactagaatt	cacgtagcgc	gtcggaaatg	atgaaaaatat			3542

FIGURE 19

catggatgca gcagatctac ggagtgcggc gcggacaaac ggcgcggtaa ttcaaagtga 3602
 gaattattagc gagagttgaa atttcaacaa aatcagccgc atttttttca aacttaattgt 3662
 attttttttc gtttttcttt ttagtaatt cggcatttgg ggctagtgtg agcattttta 3722
 agtaaaatcc tcattttccg aactccacct ttaaagggtg agtaccgaaa tttgagactt 3782
 tgctttttta ggcccaaatt ggtccaaaac taccgaattt tgtaatgaga cgttctgaaa 3842
 atttatccaa aaaatgttat ggcggttcaa agttcggcaa aataggggcc attttcagct 3902
 aaaatcaaat ttttttttcc aactttttcg gtgtcgcgaac gtctggagcc taatttttat 3962
 ttattaatca ctttttaata aatattgtag cctttgatta ggcgtttatt cgctgattta 4022
 agtacattta tggtttttgg ggcacaaata aaagtttcat tttatgcccc aaaaaccata 4082
 aatgtactta aatcagcgaa taaacgccta atcaaaggct acaatattta ttaaagagtg 4142
 atgaataaat aaaaattagg ttccagacgt tgcgacaccg aaaaagttgg aaaaaatttt 4202
 gatttttagct gaaaatgtgc cttattttgc cgcgaacttt gaaccgcat aacttttttt 4262
 gagaaagaaa ttttcagaac gtctcattac gaaattcggg agtttttaaac caatttggtt 4322
 ctaaaaagtt tcaaattcca ataaaacata ccaaagcttt gtgaaattac aataaactat 4382
 tcctaaacgt attataatcc atttcaatt cttgcag gaa gcg cat gta ttg gct 4437

Glu Ala His Val Leu Ala

210

cga atc gcc gag ctc cgt aag aac ggc tta tgg tcg aac agt cgt ctg 4485
 Arg Ile Ala Glu Leu Arg Lys Asn Gly Leu Trp Ser Asn Ser Arg Leu
 215 220 225

cca aag tgc gtc gaa cct gaa cgt aat aaa acg cat tgg gat tat cta 4533
 Pro Lys Cys Val Glu Pro Glu Arg Asn Lys Thr His Trp Asp Tyr Leu
 230 235 240

ctg gaa gag gtc aaa tgg atg gca gtt gat ttc cga acc gag acg aat 4581
 Leu Glu Glu Val Lys Trp Met Ala Val Asp Phe Arg Thr Glu Thr Asn
 245 250 255

acg aag cga aaa atc gcc aaa gtt ata gct cac gcc att gcg aaa cag 4629
 Thr Lys Arg Lys Ile Ala Lys Val Ile Ala His Ala Ile Ala Lys Gln
 260 265 270 275

cac cgc gac aag cag atc gag att gag aga gcc gcc gaa cgg gag atc 4677
 His Arg Asp Lys Gln Ile Glu Ile Glu Arg Ala Ala Glu Arg Glu Ile
 280 285 290

aag gag aag cga aaa atg tgt gca gga atc gcg aag atg gta cgg gat 4725
 Lys Glu Lys Arg Lys Met Cys Ala Gly Ile Ala Lys Met Val Arg Asp
 295 300 305

ttc tgg tcg tct acg gat aaa gtt gtg gat att cga gcg aag gaa gtt 4773
 Phe Trp Ser Ser Thr Asp Lys Val Val Asp Ile Arg Ala Lys Glu Val
 310 315 320

ctg gag tcg agg ctc agg aag gcg aga aat aag cat ttg atg ttt gta 4821
 Leu Glu Ser Arg Leu Arg Lys Ala Arg Asn Lys His Leu Met Phe Val
 325 330 335

att gga caa gtc gat gaa atg agc aat att gtg caa gaa gga ctt gtt 4869
 Ile Gly Gln Val Asp Glu Met Ser Asn Ile Val Gln Glu Gly Leu Val
 340 345 350 355

tca tcg tcg aaa tcc cca tca att gca tcg gat cga gat gat aaa gat 4917
 Ser Ser Ser Lys Ser Pro Ser Ile Ala Ser Asp Arg Asp Asp Lys Asp
 360 365 370

44/92

FIGURE 19

gaa gaa ttc aaa gca cct ggc tct gat tca gaa tct gac gat gag cag 4965
 Glu Glu Phe Lys Ala Pro Gly Ser Asp Ser Glu Ser Asp Asp Glu Gln
 375 380 385

aca att gca aac gcg gaa aag tca cag aaa aag gaa gat gtt cga cag 5013
 Thr Ile Ala Asn Ala Glu Lys Ser Gln Lys Lys Glu Asp Val Arg Gln
 390 395 400

gaa gtt gat gct ctt caa aac gag gca act gtg gat atg gat gac ttt 5061
 Glu Val Asp Ala Leu Gln Asn Glu Ala Thr Val Asp Met Asp Asp Phe
 405 410 415

ttg tac act tta ccg ccg gaa tat ctg aag gct tat ggt ctg acg cag 5109
 Leu Tyr Thr Leu Pro Pro Glu Tyr Leu Lys Ala Tyr Gly Leu Thr Gln
 420 425 430 435

gag gat ttg gag gag atg aag cgc gag aaa ttg gag gag cag aag gct 5157
 Glu Asp Leu Glu Glu Met Lys Arg Glu Lys Leu Glu Glu Gln Lys Ala
 440 445 450

cgg aag gaa gct tgt ggt gat aat gag gag aaa atg gag att gat gaa 5205
 Arg Lys Glu Ala Cys Gly Asp Asn Glu Glu Lys Met Glu Ile Asp Glu
 455 460 465

gtctcgtagga tgctcctaaa aaaattacct aaaaaaaatc gattttccct ggaaaaaatc 5265
 ctctggaaat gaccgaaaac gtcattggcg ctcgaaattt tgaaaaaaaa aaccccccaa 5325
 atttccagct aaaatctcaa attttattgc atatttttgt agttcttttg ttgtccgagg 5385
 tgctgttttc agctgaaaaat gtacctgaat ctgcaagtaa acgaccaata tatgcaataa 5445
 atgatgataa ttaattttccg atactgaaat gtgggcgaaa ttgagattt cgactgaaaa 5505
 cgtcttaaaa atcacccaaa acccggttt accgcacgaa ggtttgaaga aaatggccaa 5565
 tttttagcca aaatctcaaa tttcgtccac ttttcagtca gaaattagtt ttttgaaatt 5625
 aattaacacc ttttattgca tttttcgtc gtttattcgt tgatcgagggt gctttttcgg 5685
 tcgatgggtg cacaaattcg gtaattgtgc atccatcggc tgaaaatgct ccagaatttg 5745
 cgaatgaacg gtgaaaattt aagatttttag attgaaataa gccgtttttt agagaaaaat 5805
 ggtcgttttg agacattaaa ttcaatttaa atccctctt tattttcag agc cca tca 5863
 Ser Pro Ser
 470

tca gat gct caa aag cct tcc acc tca agc tca gat ctc acc gcc gag 5911
 Ser Asp Ala Gln Lys Pro Ser Thr Ser Ser Ser Asp Leu Thr Ala Glu
 475 480 485

cag ctt caa gat cca aca gct gaa gac ggc aac ggt gat ggt cat ggt 5959
 Gln Leu Gln Asp Pro Thr Ala Glu Asp Gly Asn Gly Asp Gly His Gly
 490 495 500

gta ctt gaa aac gtg gat tac gtg aag ctc aac agt cag gat agt gat 6007
 Val Leu Glu Asn Val Asp Tyr Val Lys Leu Asn Ser Gln Asp Ser Asp
 505 510 515

gaa cga caa caa gag ttg gcg aat atc gca gaa gaa gcg ctg aaa ttc 6055
 Glu Arg Gln Gln Glu Leu Ala Asn Ile Ala Glu Glu Ala Leu Lys Phe
 520 525 530

cag cca aaa gga tat aca ctt gag acg aca caa gtc aag acg ccc gta 6103
 Gln Pro Lys Gly Tyr Thr Leu Glu Thr Thr Gln Val Lys Thr Pro Val

45/92

FIGURE 19

535	540	545	550	
cca ttc ctg att cga gga caa ctg aga gaa tat caa atg gtt gga ttg				6151
Pro Phe Leu Ile Arg Gly Gln Leu Arg Glu Tyr Gln Met Val Gly Leu				
	555	560	565	
gat tgg atg gtt aca ctt tat gag aag aat ttg aat gga att ctt gcc				6199
Asp Trp Met Val Thr Leu Tyr Glu Lys Asn Leu Asn Gly Ile Leu Ala				
	570	575	580	
gac gag atg ggc ctg gga aag acg att caa acg att tcc ctg ctg gct				6247
Asp Glu Met Gly Leu Gly Lys Thr Ile Gln Thr Ile Ser Leu Leu Ala				
	585	590	595	
cat atg gct tgt agt gaa tcg att tgg gga cca cac ttg att gtt gtg				6295
His Met Ala Cys Ser Glu Ser Ile Trp Gly Pro His Leu Ile Val Val				
	600	605	610	
ccg acg tct gtc att ctg aat tgg gag atg gag ttc aag aaa tgg tgt				6343
Pro Thr Ser Val Ile Leu Asn Trp Glu Met Glu Phe Lys Lys Trp Cys				
	615	620	625	630
ccg gct ctg aag att ttg acg tat ttt ggt acg gcg aag gag cgt gcc				6391
Pro Ala Leu Lys Ile Leu Thr Tyr Phe Gly Thr Ala Lys Glu Arg Ala				
	635	640	645	
gag aag cgg aag gga tgg atg aag ccg aat tgt ttc cat gtg tgc atc				6439
Glu Lys Arg Lys Gly Trp Met Lys Pro Asn Cys Phe His Val Cys Ile				
	650	655	660	
aca tca tac aag acg gtt act caa gat att aga gct ttt aag cag agg				6487
Thr Ser Tyr Lys Thr Val Thr Gln Asp Ile Arg Ala Phe Lys Gln Arg				
	665	670	675	
gtgcgtagaa attttgaaga tttgcggcga atttggcgaa tttgcataat ttttttaaaa				6547
ccaattttac cgataattgc gaaatttttc aatttttatac agtggtcgga aattgctata				6607
attagtataa tttttgcaaa aatttggtact tttttcgaaa ttttgaacca ccataaaaca				6667
tttttgaaca atttttaaga ggtttaataa cgaaattcgt tcatttgaac acattttggc				6727
gatatgaatc gcccgaaaaat gtccccaat agacctaatt tcttaacaaa aatttaaaaa				6787
aaaatggccc aaaattgtct caaaatttcg aaaaaaaaac cgtaatttca gctgaaatct				6847
caaaatttgc caaattttcc gtctcacgga gatcagaaaa agttttttgc atttttttgc				6907
ggtttatttt agcgttattt cgtttaattt gatacatttt agcccaattt ttgcaaaaaat				6967
tatactaatt atagcaattt ctgacccctg acaaactttg aaattatcgg taaacttggt				7027
ataaatgggt tttttccaaa tttttaaagc gatattaaag gtggagtacc acaatttgag				7087
gctttgtttt tttttttgga cccaaattgg tccaaaacta ccgaatttcg taatgagacg				7147
ctctgaaaat ttctttctca aaaaaaaagt tacggcggtt caaagttcgc ggcaaaataa				7207
ggcccaattt cagctaaaaat caaaattttt tcccaacttc tcggtgtctc aacgcctgga				7267
acctaatttt tatatttca tcacttttta ataaatattg tgggtcttga ttgggctttt				7327
attcgttgat ttaagtacat ttatggtcag tggggcacaa aatgtaactt tttttcccaa				7387
agaccataaa tgtactttta tcaacgaata aacgccaat caaagaccac aatatttatt				7447
taaaagtaat gaataaataa taattagggt ccagacgttg cgacaccgag aagttggaaa				7507
atttttttat tttagctgaa taagggcctt attgtctcaa actttgaacc gccataactt				7567
ttttttgaga acgtctcgtt acgaaattcg gtatgttttg accaatttgg gtctaaaaaa				7627
acaaagtctc aaatttcttg ttagagattt tttaaaaatt gatatttttt ttttcag gcc				7687
				Ala

46/92

FIGURE 19

tgg cag tac cta att ctc gat gaa gct caa aat atc aaa aac tgg aag	7735
Trp Gln Tyr Leu Ile Leu Asp Glu Ala Gln Asn Ile Lys Asn Trp Lys	
680 685 690 695	
tcc caa cgt tgg cag gct ctt ctg aat gtc cgt gct cga cgt cgc ctt	7783
Ser Gln Arg Trp Gln Ala Leu Leu Asn Val Arg Ala Arg Arg Arg Leu	
700 705 710	
ctc ctg acc gga act cca ctt cag aac tct cta atg gaa ctg tgg tcg	7831
Leu Leu Thr Gly Thr Pro Leu Gln Asn Ser Leu Met Glu Leu Trp Ser	
715 720 725	
ttg atg cat ttt ttg atg cca aca ata ttc tca agt cat gat gat ttc	7879
Leu Met His Phe Leu Met Pro Thr Ile Phe Ser Ser His Asp Asp Phe	
730 735 740	
aag gat tgg ttc tcg aat ccg ttg aca ggg atg atg gaa gga aat atg	7927
Lys Asp Trp Phe Ser Asn Pro Leu Thr Gly Met Met Glu Gly Asn Met	
745 750 755	
gaa ttc aat gct cca cta atc gga cga ctt cac aaa gtg ctc cgt ccg	7975
Glu Phe Asn Ala Pro Leu Ile Gly Arg Leu His Lys Val Leu Arg Pro	
760 765 770 775	
ttt att ctg cgg cgg ctc aag aag gaa gtt gag aag cag ctg cca gag	8023
Phe Ile Leu Arg Arg Leu Lys Lys Glu Val Glu Lys Gln Leu Pro Glu	
780 785 790	
aag act gag cat att gtg aat tgt tcg ttg tca aag cgg cag aga tac	8071
Lys Thr Glu His Ile Val Asn Cys Ser Leu Ser Lys Arg Gln Arg Tyr	
795 800 805	
ctg tac gat gac ttt atg agt cgt aga tca aca aag gag aat cta aag	8119
Leu Tyr Asp Asp Phe Met Ser Arg Arg Ser Thr Lys Glu Asn Leu Lys	
810 815 820	
tct gga aat atg atg tcg gtg ctc aac att gtg atg caa ctc cga aaa	8167
Ser Gly Asn Met Met Ser Val Leu Asn Ile Val Met Gln Leu Arg Lys	
825 830 835	
tgt tgt aat cat ccg aat ctc ttc gag ccg cgg cca gtt gtt gct ccg	8215
Cys Cys Asn His Pro Asn Leu Phe Glu Pro Arg Pro Val Val Ala Pro	
840 845 850 855	
ttc gtc gtt gag aag ctt cag ctc gat gtt ccg gct cgt ctc ttt gaa	8263
Phe Val Val Glu Lys Leu Gln Leu Asp Val Pro Ala Arg Leu Phe Glu	
860 865 870	
att tcg cag caa gat ccc tcc tcc tcc tca gct agt caa att ccg gaa	8311
Ile Ser Gln Gln Asp Pro Ser Ser Ser Ser Ala Ser Gln Ile Pro Glu	
875 880 885	
att ttc aat tta tcc aaa atc ggc tat caa tct tcc gtt cga tct gca	8359
Ile Phe Asn Leu Ser Lys Ile Gly Tyr Gln Ser Ser Val Arg Ser Ala	
890 895 900	
aaa cca ctc atc gaa gag ctt gaa gca atg agc act tat ccg gag cca	8407

47/92

FIGURE 10

Lys Pro Leu Ile Glu Glu Leu Glu Ala Met Ser Thr Tyr Pro Glu Pro
 905 910 915

cga gca cca gaa gtt ggc gga ttt cgg ttc aat cgg acg gct ttt gtt 8455
 Arg Ala Pro Glu Val Gly Gly Phe Arg Phe Asn Arg Thr Ala Phe Val
 920 925 930 935

gca aag aat ccg cat acg gaa gag tcg gag gac gaa ggt gtt atg aga 8503
 Ala Lys Asn Pro His Thr Glu Glu Ser Glu Asp Glu Gly Val Met Arg
 940 945 950

agt cgt gtt ctg gtgaattttt aggaaaattg agaaaatgat ctaattgttg 8555
 Ser Arg Val Leu
 955

aatttttttaa agaatttatg ggccacaagc cgatttgccg gaaattttga tttttggcga 8615
 tttgccgaaa attttgattt ttggcgattt gccagaaatt ttgatttttg gcaattatcc 8675
 gatttgccgg aaattttgat ttttgccgat ttgccagaaa ttttgatttt tggcaattat 8735
 ccgatttgcc ggaaattttg aattttggca attttccgat ttgccggaaa ttttgatttt 8795
 tggcaatttg ccgaattgcc ggaaattttg atttttggca atttgccgaa ttgccggaaa 8855
 ttttgatttt tggggatttg ccggaaattt tgatttttgg caatttgcct atttgcctga 8915
 aattttgatt tttggcaatt tgccgatttg tcggaattt tgatttttgg caatttgcct 8975
 atttgccgga aattttgatt tttggcaatt ttccgatttg ccaaaaattt tgatttttgg 9035
 cgatttgccg atttgccgga aaaacatttt gtgagccaat tttctcgaaa tttgggcttc 9095
 aatattttca aattattcca aattttccac tgattccgaa tatctaagta aaaaaaatt 9155
 ccctgatttt atatttcagc ttaaaatcgc taattttcgc gtcagagacg acgtcatgtg 9215
 tcgatttact ggatttttaa tctttgtcgg atgctaattt ccgtttttca acgagtttcc 9275
 ttcatttcca tcggtttttg acgaagtttt ctttgaaaat atgttcttaa ggtcaattaa 9335
 acgtttttatt atcaaaaaaa actagcaaaa ttggctttta aaacacattt tcacagaaaa 9395
 ctccgacaaa aaccgacgaa aatgaaggaa acccccgtt tgaaaacaga aattagcatc 9455
 tgataaagat taaaatcccg taaatcgaca catggcgtct ggcgctctctg gcacgaaaag 9515
 tcgcgatttt aagctgacat acaaaaaaag agggatatat ttttttacga atttttcaca 9575
 tagatattcg aaatcagggg ggaaaatttg gagaaaattg agaaaatttc tcagatttcg 9635
 gattaaaaat attcaatttt tgttttctta tattaaaaaa aaattaactt ttataatttt 9695
 tcag cca aaa cca att aat gga aca gct caa cca ctt caa aat gga aat 9744
 Pro Lys Pro Ile Asn Gly Thr Ala Gln Pro Leu Gln Asn Gly Asn
 960 965 970

tca ata cca caa aat gct cca aat cgt cca caa act tca tgc att cgt 9792
 Ser Ile Pro Gln Asn Ala Pro Asn Arg Pro Gln Thr Ser Cys Ile Arg
 975 980 985

tca aaa acc gtc gta aat aca gtt cca ctg acc atc tcc acc gat cga 9840
 Ser Lys Thr Val Val Asn Thr Val Pro Leu Thr Ile Ser Thr Asp Arg
 990 995 1000

agt ggt ttt cat ttt aat atg gcc aat gtt gga aga ggt gtt gtt cgt 9888
 Ser Gly Phe His Phe Asn Met Ala Asn Val Gly Arg Gly Val Val Arg
 1005 1010 1015

ttg gat gat tca gca cgt atg agc cca ccg etc aaa cgt cag aag etc 9936
 Leu Asp Asp Ser Ala Arg Met Ser Pro Pro Leu Lys Arg Gln Lys Leu
 1020 1025 1030

acc gga act gca acg aat tgg agt gat tat gtt ccg cga cac gtt gtt 9984
 Thr Gly Thr Ala Thr Asn Trp Ser Asp Tyr Val Pro Arg His Val Val
 1035 1040 1045 1050

48/92

FIGURE 19

```

gaa aag atg gaa gaa tcg aga aaa aac cag ctg gaa att gtt cga agg 10032
Glu Lys Met Glu Glu Ser Arg Lys Asn Gln Leu Glu Ile Val Arg Arg
                1055                1060                1065

cga ttt gag atg att cgt gct ccg att att cca ctg gaa atg gtt gcg 10080
Arg Phe Glu Met Ile Arg Ala Pro Ile Ile Pro Leu Glu Met Val Ala
                1070                1075                1080

ctg gtt cga gag gaa att att gca gaa ttt cca cgt ttg gct gtg gaa 10128
Leu Val Arg Glu Glu Ile Ile Ala Glu Phe Pro Arg Leu Ala Val Glu
                1085                1090                1095

gag gac gag gtt gtg cag gag agg ctt ttg gag tat tgc gag ttg ttg 10176
Glu Asp Glu Val Val Gln Glu Arg Leu Leu Glu Tyr Cys Glu Leu Leu
                1100                1105                1110

gtg caa aggtagaatt ttgaaaatta ttactttgct tttttttaaa ccaaaattgg 10232
Val Gln
1115

cccaaaacta ccgaatttcg taatgagaca ttctgaaagc ttctcaaaaa aaaagttttg 10292
gccgctcaaa gttcgggaaa ataaggccca ttttcagctg aaatcaaaaat tttttccaac 10352
ttctcgggtg cgcaacgtct ggaactaaaa ttttgaaaaa cgagaaattt tccatttttt 10412
gcaagctgaa aaatcaaaagt ttttttttcc tcaaaattgg acaaacaaaa aaattttttt 10472
ttgaaaattg atcgaaaaaa ttcaaaattt ctataatttt tcgatttttt aaataaaact 10532
ttcatcattt ttcttccaaa ttttagtttc tcgattttta cttttttcaa aaaaaaattt 10592
tttaatacga aaaaaattca atttttagct taattctttt ttagacccaa attggtccaa 10652
aactaccgaa tttcgtaatg agacgttctg aacatttctc aaaaaaaagt tatgacggtt 10712
caaagttcgg caaaataagg cccattttca tataaaatca aatttttttt ctaacttctc 10772
ggtgtcacia cgtctggaac ttaattttta tttaattatt acttttcaat aaatattgtg 10832
gtcttttatt aggcgtttat ttgttgattt aagtacattt atggtcaagt ggggccc aaa 10892
taaaagttac attttgtgcc cacatgacca taaatgtact taaatcaacg aataaacgcc 10952
taatcaaagg ccacaattat tattaaaaag gtgtgaataa ataaaaatta ggttccagac 11012
attgtgacac cgagaagtta aaaaaaattt tgatttttagc tgaaaaatggg ccttattttg 11072
ctgaacttta aaccgctata actttttttt gagaaatttt cagaacgtct cattacgaaa 11132
ttcggtagtt ttggaccaat ttgggtctaa aaaagaatta gagctaaaat tgaattttct 11192
tcgtattaaa aatttttttt ttgaaaaaag taaaaatcga gaaaactaaa tttggaagaa 11252
aaatgatgaa aattttattt aaaaaatcga aaaattatag aaattttgat cgattttttc 11312
gatcaatttt caataaaaaa ttttttgttt gtccaatttt gagggaaaaaa aaaactttga 11372
tttttcagct tacaaaaaat ggaaagtttc tcgttttcca attttttgat gtggattttt 11432
atgagaaaaa atatataatg tcacaaaaaa tagattatta tctaaaaatc gaaaaaatta 11492
aattttccag ttttcaggaa aaaaatcggt aagaaattgt ttttccatta aagggtggagt 11552
accgaatttt gagacgtgc ttttttagac ccaaaatggg ccaaaactac cgaatttcgt 11612
aatgtagcgc tctgaaaaat tttcaaaaaa aaagtgtgga ccgctcaaag ttttgaaaaa 11672
atggcatatt tttagctaaa atctcaaaat ttggcaactt atcggtgtcg cagcggttgg 11732
aacttaattt ttatttaatt gtcattcatt aatgcatggt ttggcatttc attatgtggt 11792
atttcgttga ttgagatgct ttttgtgcct gcatcgacca aaaaaccatc tcaatcaacg 11852
aaataacaca taataaaatg ccaaaatagc cattaaagga tgataatcaa ataaaaatta 11912
agtttcaacc gctgcgacac cgctaagttg ccaaaatttg agatttttagc taaaaatggt 11972
ccatttttct aaaactttga gcggtcacaa cttttttttt gagaaatttt cagagcgtct 12032
cattacgaaa attggttaggt tcggaaccaat ttgggtctaa aaaagcagcg tctcaaaatt 12092
cgttacttca cttttaaagt tttcaattta aagtataaat tatccaatca aaaattgacg 12152
aaaaaatttt ttaaaaaatt tttcttcgga aaaaaaattt aattttaatt tttgtt aga 12211
Arg

```

49/92
FIGURE 19

ttc gga atg tac gtc gaa cca gtg ctg acc gat gct tgg cag tgt cgt	12259
Phe Gly Met Tyr Val Glu Pro Val Leu Thr Asp Ala Trp Gln Cys Arg	
1120 1125 1130	
cca tca tcg tct ggt ctt cca tca tat att cgc aac aat tta tca aat	12307
Pro Ser Ser Ser Gly Leu Pro Ser Tyr Ile Arg Asn Asn Leu Ser Asn	
1135 1140 1145	
atc gag ctg aat tct cgt tct ctt ctc ctc aac acc tcc act aat ttc	12355
Ile Glu Leu Asn Ser Arg Ser Leu Leu Leu Asn Thr Ser Thr Asn Phe	
1150 1155 1160 1165	
gat acc cga atg tcg atc tca cgt gct ctt caa ttc cca gaa ctc cgt	12403
Asp Thr Arg Met Ser Ile Ser Arg Ala Leu Gln Phe Pro Glu Leu Arg	
1170 1175 1180	
ctg atc gag tac gat tgt gga aag ctt cag acg ttg gct gtt ctg ctt	12451
Leu Ile Glu Tyr Asp Cys Gly Lys Leu Gln Thr Leu Ala Val Leu Leu	
1185 1190 1195	
cgt cag ttg tac ctg tac aag cac aga tgt ctg atc ttc acg caa atg	12499
Arg Gln Leu Tyr Leu Tyr Lys His Arg Cys Leu Ile Phe Thr Gln Met	
1200 1205 1210	
tca aag atg ctc gac gtt ctg cag acc ttc ctt tct cat cac ggt tat	12547
Ser Lys Met Leu Asp Val Leu Gln Thr Phe Leu Ser His His Gly Tyr	
1215 1220 1225	
cag tat ttc cgc ctc gac ggt acc act ggt gtc gaa caa aga cag gcg	12595
Gln Tyr Phe Arg Leu Asp Gly Thr Thr Gly Val Glu Gln Arg Gln Ala	
1230 1235 1240 1245	
atg atg gag cgg ttc aac gcg gat ccc aag gtg ttt tgc ttc att ctg	12643
Met Met Glu Arg Phe Asn Ala Asp Pro Lys Val Phe Cys Phe Ile Leu	
1250 1255 1260	
tcg acg aga tcc ggt ggt gtt gga gtc aat cta acc ggt gct gac act	12691
Ser Thr Arg Ser Gly Gly Val Gly Val Asn Leu Thr Gly Ala Asp Thr	
1265 1270 1275	
gtg atc ttc tac gat tcg gat tgg aat ccg acg atg gat gct cag gct	12739
Val Ile Phe Tyr Asp Ser Asp Trp Asn Pro Thr Met Asp Ala Gln Ala	
1280 1285 1290	
cag gat aga tgt cat cgt atc gga cag acg agg aat gtc tcg att tat	12787
Gln Asp Arg Cys His Arg Ile Gly Gln Thr Arg Asn Val Ser Ile Tyr	
1295 1300 1305	
cga ttg att tcc gag cga aca att gag gag aat att ctg aga aag gca	12835
Arg Leu Ile Ser Glu Arg Thr Ile Glu Glu Asn Ile Leu Arg Lys Ala	
1310 1315 1320 1325	
aca cag aag cgg cga ctt gga gag ttg gca att gac gag gct ggc ttc	12883
Thr Gln Lys Arg Arg Leu Gly Glu Leu Ala Ile Asp Glu Ala Gly Phe	
1330 1335 1340	
aca ccc gag ttc ttc aaa caa tct gac agt att cgg gat ctt ttt aat	12931

FIGURE 19

10 of 19

FIGURE 19

```

tcgtcgtcgt cgcagcagca gccttctcca aaaagccgct caaaaaccgg caaaaaagcc 14007
tcaaaacttc caaatctcgt ctcgctcccc gtctaagcgt aaatctcagg ctccctccctt 14067
cgatccatat gtttcgtacg caccgcacgc gctcgtctct ccccgaggatt ccccgcgtaa 14127
gagaagatca cgtgggtgcg gtagtttagg tagtggtggt ggtgggtggtg gtggtagtag 14187
atctgttgga agacctgccc gccgatcagt gaagaaagaa gaatcagatg atgatgatga 14247
ggattattgc caagaagagg aagtgaagcg aaatccggca gaaaaggtec cgccgaaaag 14307
aaaacgagtt gtgtttgtgg aacctccaga ggtgaagccg ccggagccga aaaaacgagt 14367
tggtgttcct gtcctcatcat catcatcatc agctctaact actcttcac aacaaggacc 14427
gctgatttcg ttgccaaaag ctgtgccagt tgtacctcg cccaacaac aagcaccacc 14487
acagctcatc aaaaagcacc agcagactct gatgcctgtg aagggtgctc agattagtgg 14547
tggtgggtgg ggtactccag gaccatccag tgtatcgcca ggtccatcaa tcctccgaag 14607
aacggttggt ccaggcatag gcgctggtgg tgttggaagc ctaccgcttg tcagaatgcc 14667
tggtcgccct ccatctcctg gctcgcaagc tcctgctcca ccgctgagaa gtggtgttg 14727
tccaacagct cctgcagcag ctccacgcca gttcgtcgtt ccgctcgtcg gagttcgagt 14787
tatacgcag agaactccgg tcgccaccac catggtgcaa caacaacaaa gcccgagccc 14847
gttgatgttt ccagtcgggg ttgtgcaaa gcccgggcca tctggaccac caccacctgg 14907
acctccagat cgcccaggat ttggaatcta tgagaagccg agattctcac ttggatcacg 14967
aagaagccgt ggagattcgg gcccggaaga tccggcgcca ccacagccac caccaccacc 15027
cacttctagg ccaccgccc aagcctaggc gctaggattt tccttttttt tttgttgatt 15087
tttgctcttt tttgctctc tcattgattt ataactcat tttgctttaa tatctccatt 15147
tttttgatg tgtggaattt ttttttttga aaatcgggaa aaaacgaaaa atttgaactt 15207
tttggtgatt ttcagagaaa aatccgtttt taatgaaaa aatcggaata attcagattt 15267
ttcgaaaaaa aaaaccgaga aaatttcaaa ttttcagttt ttttttcaa aaaatcgaaa 15327
aaaaaagtaa attttcagaa ttatcagcca agtttttgcg attttttgaa aaatttcaat 15387
ttttggcaat ttttgggaaa aaatcaattt ttaattcaga aaattggaaa aattaagatt 15447
tttcgaaaaa aaaaacgaag aaagtttcaa attttttagct tttttcaaaa aatcgaaaat 15507
cggaattttt ttaatttttc gaataaaaaa aatcgaaaga attccaaaac tttgcgtttt 15567
ttcttgaaat tatctgaaaa ccggaatttt ttttcaaaat tcgccatttt ttgcgaattt 15627
ttgtaatctt tttccgagaa aactcgattt tttaaatctt aataattcag atttttcgat 15687
tttcttttgt tccaaaaagt caaaaaccga acaattattt atttcaaaaa ctctaaaaat 15747
tttcaatttt ttggaaattt tcgggtataa aaaaaaccca tttttaaatc aaaaaatcgg 15807
aaatttttgt gatttttcga tttttttcac tccaaaaaaa tccacacag caaaaaataa 15867
actccgcgca tttttgagcg cacttttcaa ttttttaatt cttatcacga cgtcaaaatt 15927
cggttatttt tcacacacac acattttcct cccgagcggg tctttttttc atgagttctc 15987
ccatgttttg tttttatatt tgagacattt ttttttgttg ataagtttca acttcttctt 16047
cttcttctga ctataaacgt ttttctccat gttttttgcc tgttttctgc cgattttttg 16107
acacccaaaa ttttttttca ttttcgctcg aaaatgcacg tcggttgctc tagctttggc 16167
aagtttttaa cactgatttt ctgggttttt tttttttttg cagaattttt cagagatagg 16227
gggctcattc cagcaggggt tcccactata tttcgcattt tttccaaaaa tttttgtatt 16287
ttcaaaaatt tccaaaaaga aaggggtttt ctttaccaaa tttttctcgc cacttttggc 16347
ttaatttttg ctttagagat tcgatcgaaa aaattgcgaa agtggcgaga aatctcactg 16407
gtttgatgtt tgacccctca ctatagaaaa tttgaaaaaa aaaaaaaa aaaaaaacta 16467
gacgaaattt gtggaaatct tgctggagtt tgacgagtcg atggtggatt tttcttgaaa 16527
cgaatgaaac ggtgattttg gatcgagaa atatggcgaa aaatggtgag aaatgacgag 16587
gaggaggaag aagctgaaaa tctggaggaa caaaaattgt gtggaagtct cgggaagaaa 16647
ttagaattga aattttaaag tgttctgaga attttttgtg tgaaattttt ttaaatctgt 16707
agatcaaaaa tcaaaaaaaa aaatcagaac tattacgtgt ttatccacaa agatgagaaa 16767
aatcgccata tctggcgcgc aaatgaacct gcgggaagag acaaaactac tgtagttttt 16827
aaccaatttg tgtagattta cgagctattg cgtcatcgaa ttgaatttaa ttttcaggcg 16887
tttcacacgt ttttatattg aaatttatct atttattgaa tcaatcttaa aagaaaacac 16947
aaaaaatttt ttttaaaaaa tgcgggtcaa aattaaattc aattcgatga cgcaatagct 17007
cgtaaatcta cacaaattgg ttaaaaacta cagtgtttt gtctcttccc gcgggttcct 17067
ttgcgcgcca gatattggtga tttttctcat ctctggataa acacgtaata acatttctcg 17127
gcacaataaa tttttgctga aacaagtcgc cttctttgaa gactactgca atttcaaaa 17187
cggttttttg gttggaaagc acagtacttt ttcaaagggt cacaccttct cgaatttctc 17247
ttcgtgtcga gaccaagaat gccatttttc gattttttaa aaatcaaaaa aaaaattacc 17307

```

52/92

FIGURE 19

```

tttttaaagg tggagtaccg aaatttgaga ctttgttttt ttcggcccaa aatgggtccaa 17367
aactaccgaa tttcgtaatg agacgttctg aaaattttctc aaaaaacaac gttatggcgg 17427
tttaaagtgc agcaaaataa ggcccatgtt cagctaaaat caaaattttt tcccagcttc 17487
tcggtgtcac aacgcctgga acctaatgtt tatttattca tcactttttg ataaatattg 17547
tggcttttta ttaggcgttt attttattga ttttaagctta tttatggctt ttgtggcgtt 17607
acattttgta ccctaaaaac cataaatgta cttaaatcaa cgaataaacg cctaatacaa 17667
ggctacaata tttagtagaa agtgataaat aaataaaaat taggttccag acgttgcgac 17727
accgagaagt tggcgaaaac tttgatttta gctaaaaata agccattttc ccaaaacttt 17787
gagcgtcat aacttttttt tgagaaagaa attttcagaa tgtctcatta cgaaattcgg 17847
tagctttggg ccattttggg ccgaaaaagc aaagtctcaa atttcagcac tccaacttta 17907
gcctttacct tggtgaaatt ttttaactct tagtatactt tttttttggc cgactttttg 17967
aacacaaatt cgggtgttagt ttaaaaaaac aatcaaaact aacatattat ccagacgcga 18027
aatttttgtc ggtttttctt gcgccaaaaa gtacggtaac aggtttcggc acgatacatt 18087
tttgtaaaaa ggtgctgctc cttgaagag tgtctaataa ttttcaactt tcgtttctgt 18147
tggatttttc ttcaattttt catagatgtt ttcgatgaaa caaaaaatta acacaaaatc 18207
gtcgtgtcga gaccgaaaaa aattttgcgt ctgtgcaaca aacccggaat attaaagtag 18267
catattgatc caaattgccg atttgccgga aattttgatt ttcggcaata taccgatttg 18327
ccggaacatt tgattttctg gaatataccg atttgccgga atttttggtt ttcggaaatt 18387
tgccggaaat ttagaattcc ggcaatatgc cgatttgccg gaaattttga ttttcggcaa 18447
tatgccgatt tgccggaaat tttgattttc ggcaatatgc cgatttgccg gaacatttga 18507
tttcgggcaa tatgccgatt tgccggaaat tttgattttc ggcaatatgc cgatttgccg 18567
gaaattttga ttttcggcaa tataccgatt tgccggaaat tttgattttc ggcaatatgc 18627
cgatttgccg gaatttttgg ttttcggaaa tttgccgga atttagaatt ccggcaattt 18687
gccgatttgc cggaaatttt gatttccggc aatatgccga tttgccgga atttttggtt 18747
tcggaaattt gccggaaatt tagaattccg gcaatatgcc gatttgccg aaattttgat 18807
ttccggcaat atgccgattt gtcagaagaa atcgtttgc acccacacgt gtattgattt 18867
gatttttct ag ata aaa ttc tac gac gag ctg gac gat atc atg cca atc 18917

```

Glu Ile Lys Phe Tyr Asp Glu Leu Asp Asp Ile Met Pro Ile

1515

1520

```

tgg ctt cca cca tca cca cca gat tgc gat gcg gat ttc gac ttg aga 18965
Trp Leu Pro Pro Ser Pro Pro Asp Ser Asp Ala Asp Phe Asp Leu Arg
1525 1530 1535 1540

```

```

atg gaa gat gat tgt ctc gat ctg atg tat gaa att gaa caa atg aac 19013
Met Glu Asp Asp Cys Leu Asp Leu Met Tyr Glu Ile Glu Gln Met Asn
1545 1550 1555

```

```

gag gct cgc cta cca caa gtt tgt cat gaa atg aga cgt ccg ttg gct 19061
Glu Ala Arg Leu Pro Gln Val Cys His Glu Met Arg Arg Pro Leu Ala
1560 1565 1570

```

```

gaa aaa cag cag aaa cag aac acg ttg aat gcg ttt aa tggtaatat 19109
Glu Lys Gln Gln Lys Gln Asn Thr Leu Asn Ala Phe Lys
1575 1580 1585

```

```

ttcaaaaaaa aatttttttg aaaaaattca attaaattcg attttgagca atttttatcg 19169
tgaagattgc ataattttga gattttgctc caagattttt gttaaattga aaaaaagaga 19229
tgtgcgcctt tatggagtac tgtagttttg aaaattgaaa ttacagtact ctgtttaaag 19289
gcgcacacat gtattacgta gcgaaaagaa aagtacagta attagttaaa taagactact 19349
gtagcgcttg tgcgatttta cgggctctga attttatatg aatttttgaa aactagaaac 19409
atctcaaat gcataaaatt accatttgaa cctcccgcga agtgattttg ttcgacgggg 19469
cgcgcttgca cgttttctat ttttaatttaa ttcaattttt tttgcttaat tctcaccgat 19529
ttttcatggt ttcagtttga ttttgatgga aatttgagaa caatatcaac ataaatgctt 19589
ttcaatcgaa aatgtgcatt tatattgaca ttttctccga atttccatca aaattaaact 19649
gaaaacacga aaaatcgggtg agaattaagc gaaaaaattg agttaaatga aaatagaaaa 19709
cgtgcaagcg cgctccatcg aacaaaatca attggcggga ggttcaaatt ggaattgtat 19769

```

FIGURE 19

gcaatttttca aaaggtcgta taaaattttg aagaaagcaa attaaattta aaaaatcgag 19829
 ctcgtaaatc gacacaggcg ctaatttttca aaaaaataaa atgacaccca aaaaatcata 19889
 agaaaatcat aaataaatat tacgggaaca caaaactcag agaacccgta ttgcacaaca 19949
 tatttgacgc gcaaaatatg aaatatctcg tagcgaaaag aaaactaccg taattttaaaa 20009
 acattttaa at gactactgta gcgcttgtgt cgatttacga gatctcgatt ttctaaataa 20069
 atttttttaa aaatgatgtc agcgatatc catttgactt tgtttcttcg tattattttc 20129
 tcatttttgc ttgattttat ttaattttat aattttattt aaaatcaagc aaaaacgaga 20189
 aaataatacg aagaaacgga gttaaatgga atatcgctga cataatttaa aaaaaaatt 20249
 taattagaaa atcgagatcc cgtaaatcga cacaagtagt catagtacag tagtcattta 20309
 actaattact gtacttttct ttctgctgcg agatatttca tatttttatt catattttta 20369
 tttattttca tttttttata tatatatata tatatatatt tcttggcgtt ctaatgcagt 20429
 ttctctcaat taattcc a gac att cta tcg gca aaa gaa aag gaa tcg gtg 20480
 Asp Ile Leu Ser Ala Lys Glu Lys Glu Ser Val
 1590 1595

tac gat gcg gtc aac aag tgc ctt caa atg cca caa tcc gaa gcg atc 20528
 Tyr Asp Ala Val Asn Lys Cys Leu Gln Met Pro Gln Ser Glu Ala Ile
 1600 1605 1610

aca gca gaa tct gca gcg tct cca gca tac acg gaa cac tca tca ttc 20576
 Thr Ala Glu Ser Ala Ala Ser Pro Ala Tyr Thr Glu His Ser Ser Phe
 1615 1620 1625

tcg atg gat gat aca agc cag gat gcg aag att gag cca agt ttg act 20624
 Ser Met Asp Asp Thr Ser Gln Asp Ala Lys Ile Glu Pro Ser Leu Thr
 1630 1635 1640

gaa aat caa caa ccc acc acc acc gcc act act act act aca gta ccc 20672
 Glu Asn Gln Gln Pro Thr Thr Thr Ala Thr Thr Thr Thr Val Pro
 1645 1650 1655 1660

caa caa caa caa caa cag cag cag caa aaa tcg tcg aaa aag aag aga 20720
 Gln Gln Gln Gln Gln Gln Gln Gln Gln Lys Ser Ser Lys Lys Lys Arg
 1665 1670 1675

aat gat aat cga a cggtacggag gttactagcg aacaatttca agaaattttg 20773
 Asn Asp Asn Arg
 1680

aatttgtgaa aattcaattc cggcaatttt tcgatttgcc ggaactttta attttcgccg 20833
 aattgtcaat ttgccggaaa ttttgatttc cgccgaattg tcgatttgcc ggaacttttc 20893
 attttcggca aattttcgat ttgccggaac ttttaatttt tgacaaattg tcgatgtgcc 20953
 ggaaattttg attttcgaca atttgctgat ttgccggaaa tttcaatccc aacaattttc 21013
 cgatttgccg gaaatttcaa tccaacaat tttccgattt gccggaaatt tcaatcccaa 21073
 caattttccg atttgccgga aatttcaatc ccaacaattt tccgatttgc cggaaatttc 21133
 aatcccagca attttccgat ttgccggaaa tttcaattcc ggcaattttt cgatttgccg 21193
 gaacttttca ttttcggcaa agtgtcgatt tgccggaact tttcattttc gccgaattgt 21253
 cgatttgccc gaacttttaa tttttgacaa attgtcgatt tgctggaaat tttgattttc 21313
 gacaatttgc caatttgccg gaacttttaa tttttgacaa attgtcgatt tgccggaaat 21373
 tttgattttc gacaatttgc caatttgccg gaacttttca tttttgccaa attgtcgatt 21433
 tgccggaaat ttttaattccg gcaattttgc gatttgccgg aaatttcaat tccggcaatt 21493
 taaaaacact aaaaacacaa aattttcggt ttcccggtt ttcgatgttt cagcttttct 21553
 caaaaaaftg cgattttccg aaaaatcgaa acaattttcg ggggttaaac cgggaaattc 21613
 ctaaattcct atttaaaaga attgaaaaaa aactctcaaa attcc ag gct caa aat 21669
 Lys Ala Gln Asn

54/92

FIGURE 19

cga aca gct gaa aat ggt gtg aaa cga gcg aca act cca cca cca tca	21717
Arg Thr Ala Glu Asn Gly Val Lys Arg Ala Thr Thr Pro Pro Pro Ser	
1685 1690 1695 1700	
tgg cgt gaa gag cca gat tat gat gga gcc gaa tgg aat ata gtt gaa	21765
Trp Arg Glu Glu Pro Asp Tyr Asp Gly Ala Glu Trp Asn Ile Val Glu	
1705 1710 1715	
gat tat gca cta ctt caa gca gtt caa gtc gaa ttt gca aat gct cat	21813
Asp Tyr Ala Leu Leu Gln Ala Val Gln Val Glu Phe Ala Asn Ala His	
1720 1725 1730	
tta gtc gaa aaa tcg gcg aat gag gga atg gtg ttg aac tgg gaa ttc	21861
Leu Val Glu Lys Ser Ala Asn Glu Gly Met Val Leu Asn Trp Glu Phe	
1735 1740 1745	
gtg tcg aat gcc gtt aat aag cag aca aga ttt ttc cgc tcg gcc cgt	21909
Val Ser Asn Ala Val Asn Lys Gln Thr Arg Phe Phe Arg Ser Ala Arg	
1750 1755 1760	
caa tgc tca att cga tat caa atg ttt gtt cgg cca aaa gag ctc gga	21957
Gln Cys Ser Ile Arg Tyr Gln Met Phe Val Arg Pro Lys Glu Leu Gly	
1765 1770 1775 1780	
cag ttg gtg gct tct gat ccg att tcc aag aaa acg atg aaa gtc gac	22005
Gln Leu Val Ala Ser Asp Pro Ile Ser Lys Lys Thr Met Lys Val Asp	
1785 1790 1795	
cta tcg cat act gaa tta tct cat ttg aga aaa gga cga atg act acg	22053
Leu Ser His Thr Glu Leu Ser His Leu Arg Lys Gly Arg Met Thr Thr	
1800 1805 1810	
gag agc caa tat gct cat gat tat gga ata ttg act gat aag aaa cat	22101
Glu Ser Gln Tyr Ala His Asp Tyr Gly Ile Leu Thr Asp Lys Lys His	
1815 1820 1825	
gtg aat aga ttt aaa agt gtt cga gtg gcg gca aca cgg aga cct gtt	22149
Val Asn Arg Phe Lys Ser Val Arg Val Ala Ala Thr Arg Arg Pro Val	
1830 1835 1840	
cag ttt tgg aga ggc cct aaa ggt aga gga gga tgg ctt cat aat agt	22197
Gln Phe Trp Arg Gly Pro Lys Gly Arg Gly Gly Trp Leu His Asn Ser	
1845 1850 1855 1860	
cac tgc aac ttt ttc ctc acg agg gac gag aaa aag tgg ttt cta ggc	22245
His Cys Asn Phe Phe Leu Thr Arg Asp Glu Lys Lys Trp Phe Leu Gly	
1865 1870 1875	
cat ggc cga ggt gcc gac aag ttt ca gcggccattt atcttgcttt	22291
His Gly Arg Gly Ala Asp Lys Phe	
1880	
gttttccgcc cgttttcttt cgtttttcac cgattttttt cgttttttct taataaaaact	22351
gataaataaa tatttttttgc agatgctaaa aaaatttcca agtaaaaaaa tcatgtattc	22411
agtgggcatg cagcggtgaa agtgggcatt gtaatatgat ggattacggg tatacaaaac	22471
ctaaactttt tctgaaacat gatacatgtg ctgcttaaat gctgagacta cctgattttc	22531
ataacgagac cgctgaaaaa gttttgaggt ttccaaaatt caactttttt aatgaaaaag	22591

FIGURE 19

```

tcgagatttt cgcacaaaaa gttgaatttt gaaaacctca aaactttttc agcgggtctcg 22651
ttatgaaaat caggtaattt cagcatctaa gcatcatatg tatcatgttt cagaaaagtt 22711
taggttttgt attcccgtaa tccatctatt tacattgacc actttcaccg ctgcttgccc 22771
actgaataca taattttttc acttggaat tgttttagca tctggaaaaa gtattttatt 22831
atcagtttta ataagaaaaa acgggaaaaa gctgtgaaaa acaaaagaaa acaggcggaa 22891
aacaaagcaa gataaatggc cgtgaaact tgtcggcccc tcggccatgg cctagaaacc 22951
acttttcttc gtccctcgtg aggaaaaagt tgcagtgata gtctaaaatt cggaggaatt 23011
ttttaaaatt ggaaaaaatt gtttaatttt tttttcttgg aaattggaaa atcacaattt 23071
ttcgattttt gtttgttaaa aaaaaaaaga aaattggcat aataaaacat ttcttttttt 23131
tttgaaaatt gggaacttct taatatcaga ttttttaagt aaattttttt tgattttccg 23191
gaaattcgga aaacctgaaa attttcaaca tttcaaaata aaaatttccg tttttttttt 23251
ctgaaaatct ccaacaaaaa aagggtcaaat cgtcagaatt attgttgga gtggcggttt 23311
ttcacgatta gagttcagta ttttttcttc tgaatttcaa atttgaaaa aaatcgaata 23371
aactgtagaa aatgataga aaattaacaa aaattctgat taaaggtaaa gggaaaatag 23431
accgtaatga ccgaatataa ctgttgaaaa tatcaacaaa aaaaattctg aattttttgt 23491
gactttttca atttttcaag aataaaaaaa acgaccgaat aaaatatttg aattcccgcg 23551
caaatgagtg actggttctg gccaatttac agtcttttta taaaagaaaa aatctagaaa 23611
aaccggcgaa tttagccaga aaacgcaaaa aattaaaaat gacgtcactc atttgcgcg 23671
ggaatacaaa ttttaattag ccgtttcttt gatttttgaa aaattgaaaa aaccattaaa 23731
aaatttagaa attttttga attttttaca gttttttatt cggtcattat ggggttattc 23791
aagtagtgtc ggaataattaa aaagtgtaga aaaattacgt cacaactctg tattcaagta 23851
tataaaaaa tgtattttaa tacattttgc tacattactt gaataacccc attaggggtt 23911
attttcttta gagcaaaaaa aaacatgttt ggctctactc cacctttaaa tgaaaaaatc 23971
gacaaattgt gattttgcaa tttccagaaa aaaaagaaaa aagttgcttt ttggaaaaaa 24031
ccaaaaaag ccatttgaaa aatttttatt tccaaaaaaa attattttgc agctctagaa 24091
tctcgaaatc tgcaatctct aaacggcgga atgccaccac gacacgagtc gagactcgcc 24151
gaattcgacg taaaaacca tattcgccctg gacgccgagg acattgtcac aatgtccgac 24211
gagtcgattg tcgcctatga agcagcaag aagaagctac tggccagtcg tcaaacaaaa 24271
ccctcaccac gtcaagatgt ccgattccat acgctggttc ttcggccgta taccgtacct 24331
gtgacaactg agtactcggc tgcaccttct cgtcgtgaaa tgcgcacgc tgttccaccg 24391
cttcagcctt cggctttatc tacgatttcc tcagttgctg ctgctgccac gtctgggcca 24451
cttaccatcaa ttcagcattt gcagtcgtcg tctacgggct tgggatctca gcaaaatttg 24511
caaaattcgc ataattctga gcaaagaaat aatgtgcaaa atatgcatca aaatcaatat 24571
aattcaagtc aaaatccgcc aataacctat cgacaaatcg gagcagcatc atcacacca 24631
catgatcaag gatctcaggg gcctggggga aaaccacaag cctatcacct ggtgcaacag 24691
ggatcacagc aacagcagca gcagcagcag caggcgacgt tacagcgaag aaatgcggcg 24751
gcggcgcgag ggtcgaatgt gcagtttatt cagcagcagc agcagcagca gcaatcgggt 24811
aaaaattgta tggatttata ggaaattata tgaatttgcg cggggatagc cccggcgaaa 24871
aacgggaaaa agcgacaatt taiaaaaaaa tcgtgtgaaa atctcaattt tttaaat 24931
tgaaagtaat tttttattga aaaaagtga atttaggcat tcatccagag cagggtggtg 24991
accataaaaa atttttggac caaaaaccaa aaacacaaaa attgaaattt ccgaaaaatc 25051
aacttaagca tcaaaaattt tttgtttttt tttttgtttt ttggtttttt ttggtttttt 25111
gacgaaaaaa cgattttttg gttttttggt ttttcgagac caaaaaaacc aaaaaatcca 25171
aaaaaatgtt tgccgtgtct agtctcgacc tagacacggc aaacattttt tttttttgga 25231
ttttttggtt tttttggtcc cgaaaaacca aaaaaaccaa aaaatcgatt tttcgtcaaa 25291
atacaaaaaa aaacaaaaga attcccagcc cctttcgcca aaattgccgg atattttcaa 25351
acctcaaaaa aaattttata aggtggacta catcctgtgg ggaaattgct ttaaaacatg 25411
cctatgggct cacaatgacc gaatatcatg attaaaaaat tcaacaaaaa aattactaga 25471
ttttatgtga tttttgaaa attaaaaaaa tctcagtttt caacctaat cctatttgaa 25531
tttccgccaa tttgatttgt tcgatggagc gcgcttgctt tatttttttt tattcattga 25591
ttttattttt attagcatta tttcactgat tttcttcatt ttttggtgtg ttttggtggt 25651
aattgaaatg aaaaaaaaga agataaatgc agaaagtctg ttaaaagggtc attgaaatg 25711
cttaaaacgg caacaagctt gaaatttgta tattttacac agttttacgc atttttcaatg 25771
actttttaac aaactttccg catttatctt gtttttttca gttcaatttc cattaaaaaa 25831
cacacaaaaa aaaaatgaaga aaatcagtga aataagggtta ataaataaaa taaatgaata 25891
aaaatgatgc aagcgcgctc caacgaacga attcaattgg cggaatttca aatatggaat 25951
taggtgaaaa ctgagatttt tttttcaatt ttcaaaaaat catataaaat ctagaacct 26011

```

56/92
FIGURE 19

```

tttttgaatt ttttaatcat gatattcggg cattgtgacc ccataggcgt gttttaaagc 26071
aatttcccca caggggtgtag tccacctttg acgagggttg aaaaatgtccg gcaatttttg 26131
cgaaattgcc ggaaacttga gatttttcag tgaaaaattc caaatttcat gtggaaaact 26191
gtttttttgt tttttgaaa atgcaacaaa aaaaactatt tggcgcgaaa acgcggatag 26251
ttttgccaat tttcaaggat tttccgctat ttttaatgtt tttatgccga attttacttt 26311
aaaaaatcat aattattcgg aaaatgctcg aagagcattt ccaattgtct gtggagcgcg 26371
tttgactaat cagataatat tccaggcggg caaggacaaa gcttcgttg catgggctcg 26431
cagagctcat caaatgatgg acaagggtgga gcacgcagcg tcggaggagg aggaggagga 26491
tcacaacagc ctcaccagca gcagcagcag cagccacaac aaagaatata gtacattcca 26551
caagttaccg gtacgcgaaa taacgggtgga ggtggtggaa gaggaggcta cggtagtaca 26611
ctggtcatgc caagaggagg acgtgttgct aggggttgga gaaatacaaaa atcgcgaaaa 26671
aacggcattt ccggcttccc gaccaatcag cgattgtct cgcaccttt cggaccaatc 26731
cgctgaccga ggcatttgat tgggttgaaa ttgggcggag cagcgaattg ctgatgcgaa 26791
atacgggaag ttctcatttt gatggaaatt ctgcataaatt ctttaaaaaa aacaaaatct 26851
tctcaaatcc ggaaaaaatc acaaaggaaa tcgaagaaaa tcgcgatttt tgattccccg 26911
accaatcagc gatttgctcc gccacctttt gaaccaatca gcgttcgagg catttgattg 26971
gttcaaaaact gggcggagca gcgagttgct gattggattt ttcagttttt aaatttttaa 27031
agcttttttt aacggaaaaa ttcgagaaaa ccatagattt tgatgagaaa tgatgaaaat 27091
tttcatgaaa aaatggaaaa atgattggaa attaatcaaa aaatcttgaa aaaaaatttt 27151
ttttcagaga aaatgcttca tttttggctc tgaaacgcct cttttttatt tgtgcctccc 27211
cgaccaatca gcaatttgct ccgccactt ttgaaccaat cagcgaccga gcgatccgat 27271
tggtttgaaa ttgggcggag ctaaaatgat ttttaaaaaa ttcccgattt gtttaattca 27331
gaaattttaga aaaaagaaat atagaaaaaa aatagaaaaa aattaaaaaa aaaaaaaca 27391
aaaaatcgga aacgtcggaa aatattacga aaaaaatttt ttttaattgat tttttttcga 27451
aaaaaactaa aattttaacc aaaaattcaa agaaaaaatt tgtttttgat tttttttcgc 27511
aaaaaaaaaa aatttttaac caaaaattca aaaaaaaaat gtttttctg atttttttcc 27571
aaaaaaacta aatttttgac caaaaattca gcaaaaaaaa aattttttta ttgatttttt 27631
tttcgaaaaa aaataaaatt ttaacaaaaa attcaaaaaa aaaatttttt attgactttt 27691
ttcgaaaaaa actcaaattt taaccaaaaa ttcaaaaaaa aaaatttttt ttttgatttt 27751
ttccgaaaaa aactaaaatt ttaacaaaaa attcaaaaaa aaaatgtttt tcttgatttt 27811
tttccaaaaa aactaaaatt ttgacaaaaa attcagcaaa aaaaaaattt tttaattgat 27871
tttttttcgc aaaaaaata aaattttaac caaaaattca aaaaaaaaat tttttattga 27931
ctttttcga aaaaaactca aaattttaac aaaaattcaa aaaaaaaaat ttttttttg 27991
attttttcgc aaaaaacta aaattttaac caaaaattca aaaaaaaaat tttttattga 28051
ttttttccaa aaaaaactaa attttgacca aaaattcagc aaaaaaaaat ttttttaatt 28111
gatttttttt cgaaaaaac taaaaatttg accaaaaatt caacaaaaaa aaaatttttt 28171
attgattttt ttcgaaaaaa actaaaattt tgacaaaaaa ttcaacaaaa aaaaattttt 28231
ccagccagcg ggaactctac caggcgggtg acgtctgtat gtcgatacata accgtcatcc 28291
atatccaatg tcgtcaaatt ttgtgccagt acgtgttcta ccagccacgc aacaaggaca 28351
acaacgaatg atgacaggac aacgtcgtcc ggctccagcg cccggtactg tcgccgcaat 28411
ggtgttgccg aatcgaggag ctggtggaat tccgcaaatg cgcagtttg agtgagtttt 28471
gcacggaaat tggacgattt tcagcgaaat tttcgggaaa aatggctatt ttgtgtttga 28531
aattgcgaaa tttcacgatt tcgtcttaaa tacgggtgca acctaccca tgacggtttg 28591
atctacaaaa aacgcgggaa tttttcacac aaaaatatgt gagacgtctg cacgttctta 28651
accaatcggg tgaaaaactc gccgcatttt tgtagatcta cggtagatca ctgcagattt 28711
taagagagaa aaataaataa ataatccac aagggtttta aaattttttt ttcaatcgta 28771
aaaaatagcg aaaaattggt tttcgcgtcg agaccctacg cacatttttt tgcaattttc 28831
gcttcaaaat tacggtaccg ggtctcgaca cgacattttt attgtgtaaa atacacaatt 28891
ttttggaatt ttcacgatt cgaatttaaa tatttttaaa tgatttaatt aattcttaac 28951
gaaaaaaaaa agtttcgaaa ctgcagtact ctttaaaggc gcacacatgt atgtatttat 29011
aaaaaatgtc gtgtcaagac cgtacttttg gctcacaatt tgcaaaaatat tgcggaattt 29071
tttttaattt tagataaaaa aaaacatgaa aaatctatgg aaactaaact tataatttaa 29131
aaaaaatatt ttttaagggt gactacgctc agtggggaaa ttgcttttaa acacgcctat 29191
gaggcccaa tgactgaata tcatgattaa aacaatacaa aaaaattttc tagattttat 29251
atgatttttt gaaaattgga aaatcacag ttttcaccta attcttttg aatttccgcc 29311
aattggatta gttcgggtgga gcgcgcttac attattttta attatttatt ttattttatt 29371
tcgttatttg actgattttt ttcatttttt gtgtgttttc ctcggaaaaa ggaagaaata 29431

```

57/92
FIGURE 19

```

aacaagacaa atgcaaaatg tttgttaaaa agtaattgaa aatgcgtaaa accttgatat 29491
tctgagttcc gacgacaaca agcctgaaat tagtatattt cacagttttt ctcattttca 29551
attacttttt aacaaacatt ttgcatttgt cttgtgtatt tcttccattt tccgaggaaa 29611
aaacatagaa aatgaagaaa atcggtcaaa taacgagaat aaataaaaatt aattttaaaa 29671
aagatgcaag tgcgctccac cgaacaaatc caattggcgg aaattcaaat atggaattag 29731
gggaaaaactg tgatttttcc cattttcaaa aaatcatata aaatttgga aatttttttg 29791
aattttttta atcatgatat tcggtcattg gcgccccata ggcgtgtttt aaagcaattt 29851
ccccactgag cgtagtccac atttaatttt ccaaaacagc acatgcta at cctccaagtt 29911
attccagacg aggcagttac accggcgggt gtggtcagca acgaatcaac gtgatggttc 29971
aaccacaaca aatgcgcagc aacaatggcg gtggagtcgg tggcacaagg ggcctccagg 30031
gtgggtccagg aggtccgcaa ggaattcgtc ggccactcgt cggacggcca ctacaacgag 30091
gagtcgataa tcaggcggcg acggttgctc aggtcgttgt tgctccgccc caaggaatgc 30151
agcaggcatc acaaggacca cccgtacttc atatgcagag agcggtttcc atgcaaatgc 30211
cgacgagtca tcatcatcaa ggccaacagc aggtcctctc gcagagctca cagcaggctt 30271
cgcaacaggc tcccacatcg gattctggga cgagtgtctc gccacgacaa gcaccaccac 30331
cacaaaaact gaattttccc ctattatcct attttacccc ccaaaaactct attaattaaa 30391
taattttcctt cctatttttt tcttcgttgt aagattattt gtcccccaac caagggtgtc 30451
ggtttttctga ttttctgacg tttttcaaaa aaatttcgat ttttcaaaaa attagcttca 30511
tattttggct attactctgc tttttagaag aaatttgtat gtttttctt gaaaaataaa 30571
gcaaaattag atttaaaaaa atcatattt tatggttaat tttctgaaca tatttttcaa 30631
ttttcgattt tcacagaaaa acatcgaaga atcgacaaaa tcgaaaaata tgttccgaaa 30691
attaaccata aaatatgatt ttttttaaaa tctaattgtg atttatatta taagaaaaaa 30751
catacaaat tcttctaaaa agcagagtaa tagccaaaa atgaagctaa ttttgaaaa 30811
aacgaaaaat tttcgattt ccaagaatc gaaaaatcga aaaatgacac ccttgcccc 30871
aactatctct gtatattatt catctattat tgattgtttc ttttgttcc tcgaaatttt 30931
ttgaaattaa agttctcttc cccacccgga tttccgttgc tttattaatc gcgattgatt 30991
aattgttttt ccataaatcc ccaactattt atctctgtat attattcatt tatattattt 31051
atcttttctc tgtgtcgatt tacggtatct ccgggcccgt tgattttgaa ttctctctc 31111
aaataaaaatt gttttctatc taacatttga tacgtgtttt tctgattttt ttgtatatat 31171
attttccatg tatatatatt ttttctttt ttctttgtc caactttatt ttaaataatg 31231
cttttttctc aagagatttt ttaaaaaatc gatttttttt aaagccagga attctgaaga 31291
atcgaaaaaa atggaactat ttttcaata atgagaaagt tttttttttt caagaaaaaa 31351
ataataaaat tctgattttt ttaataaaaa ttaataaagt ttttgaagat ttctattgaa 31411
aacatctaaa ctattcgatt ttgattttt aattttgaaa atagaatttt ttaatatatt 31471
tttttcaaat cgttaaaaaa agaatgccgg aattttttta aaattcttta aatttagaaa 31531
taatcgaaa attttcgatt ataaaacggt gtataaaacg aaaaaaagt gattttgatg 31591
aaagaaaaaa tttttttgta gtttttttca gaaaaaaatt actttttatt ctccattttt 31651
tgttgttgaa tttttgagaa aaaactcatt ttgaaaaaat cgaatttttt atattttttc 31711
taatcgtaaa aaaaattttt aaaaatgaat ccggtaattt tttaaaaaat aatattaatc 31771
tatagttttg tagttaaaaa aatgtttcac ataaaaatct aaaaattttt gatttttaaa 31831
taaaaaaaa tcgaattttt taaaattttt ttcaaactcg aaaaaaagaa acaataaaca 31891
aaagaattcc ggaaaaaaat tatattatga ttataaattt atagttcttt acttttttaa 31951
aagattttat tttaaaaatt ctaaaatgat cgattttttg ttttttaaaa taatcaaaaa 32011
tgtttgattt tttttaaacg tgaaaaaaat gcaaaagaaa tgaaatccgg caaaaattgt 32071
aatataatta taaatctata cttttgtggg tttttccaat atttctataa attcttgatt 32131
tttaaaaaa tcaaaagttt tgatttttaa ttttgaaaa attgaatttt tgatatattt 32191
tctaactgta aaaaaatttt taaaaaatc gaaagcggat ttttttctgc tttttgttt 32251
tttttttgaa aaccggaaaa aataccaaaa attgatagtt tcgaccactc tggctagact 32311
acaaaaattg aatttttttt ttcaatttga gaatggcgt ggtctcatca gtagctagcc 32371
attctctttt tttttcaatt ttaagaaaa aagtctctaa aattttgaaa aaatcgattt 32431
tttttactta ctttgatact tttttatat cttttcaaat cttaaaaaac aattttaaaa 32491
attgaattcc ggaaattttt ttaataata taaatctata gttttttagt ttttaaaaaa 32551
tatattttta taaaaatcta aaaagttcgg cttttgactt ttgaaataat cgaaaatggt 32611
tgtttttaaa tttgaaaaaa tataaaaaat tcgatttttt caagataaaa agcgaatttt 32671
tttgaatttt tttcaaatcg taaaaaatgt ctgtagtttt tttaaagact ctcataaaaa 32731
tctgaaatgt tcgatttttt atttttaaaa taatttttaa aaaattttta ttttttttat 32791
cgtgcgaatt ttttaccac tataatttgg aataattttc aqatctcaa aatatccac 32851

```

FIGURE 19

aatcgcgcaa	atâtgccagg	aagcaatgaa	gattggataa	agaaggagggt	cgaggaccag	32911
gacaccaacg	ccaacagctc	gagctccagc	atagccgtct	cgcgtcagct	cgaaggggaat	32971
tctgctgttc	ctgacgccat	cgaccttctg	tcttctcaaa	tcaaaagaga	agttgaagag	33031
gaggatgatk	gcaacgatga	gactggaccc	cgttcggagc	ccgtggatgt	taagccgtct	33091
ccaaaacgcc	caacgaagag	gtcagccgag	acctggacga	cggtcggcg	ccaagcaaga	33151
aacggtctac	ggcgggagac	ggttcaactc	atcgattcgc	gtatgtgaat	gttggagtcc	33211
gccatccata	cgatccacgc	catcttgtca	tggaaaacttc	attgaatgaa	attaggtaa	33271
gaattattga	aaataattat	tatatattca	tttttaattca	attttttttt	tcagaatcga	33331
agatttcgaa	ataatccagt	atcttccgat	gcccttcagg	acttcgattc	ccatgaagct	33391
agtgatcttc	gcagtgagaa	gtgaagaatc	tgccgagaag	atccgctcgt	taatcgatcc	33451
ttcgatgttg	atcgcggtt	ttggtggcgg	aaccgaaact	caaaaattct	tgtggagcga	33511
gctgacggtg	gaggatttcg	tcaaggcaca	cataatggcc	agcaggtaag	ctttcgaaca	33571
tacttaattt	tttaaaaact	aaaattcagc	gcaaccgatg	acgtgccata	tgaggcagcc	33631
atggcggatc	gagaatcgct	caaacaagct	gtaaatgatg	ccagctctct	gaaaggcttg	33691
aaggaggtaa	taatttagaa	atgacagaaa	atgaaccgtg	atgacgaaat	acatctgtaa	33751
aaaaattata	aaaaattcta	agctccgttt	ttaatttttt	ttttcagtta	tattctgtca	33811
tagcggccta	tttctctgga	aaaaaaaaatc	caaaatagcc	tcaaattcgg	aattatgctt	33871
cgattttttt	tctgcggtag	tctgaattt	aagacgattt	tgaatttttg	tagctgcctt	33931
tcgccacaat	tacgttaaac	atttcagagc	atgtcgaaag	ctggatggag	gacgtgagt	33991
aagatgcgga	aagatctcaa	tggagcctga	tgatccctt	cccagcacac	aagacagttt	34051
taattttgtg	tctgtatagt	tttatattaa	gttttgatga	taatgaattt	ttttacggtt	34111
ttatccatca	cttggctcga	ttgaagctcc	tattgtgcag	cacacacggc	gtgtaaaatta	34171
gtgcatctaa	cctaggaaat	gcgatttcta	ggccatggcc	gaggatccga	ctagatcttt	34231
tttgatggtg	tttgtagaga	gttaaatttc	attttgagg	gaaattgaag	gaaattgaaa	34291
gagaaattaa	tttaataata	ttaatttgat	ttaaattgacc	agaacaaaac	aaataaaactg	34351
aatgacaagc	caatcgatat	tcgtccagac	tgggatgatg	ttatatgaac	tctttcacct	34411
gaaacattta	agttttttta	ataaaaagagc	aagcgcgctc	aaacgcgaaa	acgctcgatc	34471
cacttaatat	ggattttgtg	ccgattcatt	tatttcaagc	tatgctcggt	tttttctgtt	34531
atgtttcatt	aaaaagaccg	aaaacataac	aaaaagtgcc	tgaaaacgaa	aaaaaacccg	34591
cgacattaat	tgaaaaattc	aaaactacaa	tttcgccgcg	aaaacccaac	gagacccaaa	34651
gtttcagcgc	ggagcgttcc	cacttggcgc	tggagcgcgc	ttgtatataa	aaggacttaa	34711
ttttttaaaa	tacttaccgc	agttacttcc	aatgtatgtc	aaattcactc	gattctccat	34771
tgcagggtta	ctaaaatatg	ctccaaatag	ttggcaaggc	gttgacttga	ataaatcggg	34831
atggttatct	tggatgattg	cagttcgatt	tccttttgta	attatgttct	aaaaagtcac	34891
tgtaatcatt	taaaagtgga	gtagcgcag	tggggatttt	gtctaaatgc	acttattatg	34951
atccaaaaca	accgaatatc	atcataaaac	actccaaaaa	gtttagtttt	ttcataattt	35011
cctgtcaaaag	ttttggcaaa	ttggcaaaat	tttgaaaaat	gcgagctttt	gaggtaattt	35071
aaggaaatgt	cgcattgttc	gacccttaca	attatttaat	acagataatt	taaacaaaat	35131
taaaacataa	aaatgtagaa	attttttttg	ttttggctga	ttttcaaaat	tatgagtggc	35191
aaaaactgag	taattgccac	tttttgacag	taaaataaaa	atgttcaaaa	ttttttgaaa	35251
cgttttatca	tgatatttgg	ccattatggg	agcaaatgag	tggtttatct	attttttcac	35311
tggcgcctact	ccacctttaa	gcattgtctgc	ctcaccataa	tcccatttaa	tccaacggtt	35371
cttagatttg	gattcgaata	tatttgaatg	actggaaaaat	atgttacggt	accattcaat	35431
gcaccaatat	aagtcatttg	atcgagaaaa	ttcaaatcgg	tgagatttgt	gtttctgata	35491
gtcaatgttc	cgaataaaaa	ttgtaacact	cctaatttgg	aaacataatt	ttcatcttca	35551
tggcttatta	atagatctcc	aaggatatac	atacatgtat	ctgatagttt	gctcattgat	35611
tcaaatgtgc	aataaaaatga	cgcattccaat	ggaccaggat	ctttgcaaag	tttcgcttca	35671
atgttttcag	tagaaattcc	aaggttcaat	agggcaacta	tctcagtaat	ggtgacacaa	35731
aaatcaggat	gaaggttttc	aaaattgaag	tattgccttt	tattgtatgt	actgtattgt	35791
atcatactgg	tttgctcaac	tgtatctata	actttctgaa	attttatgtc	attattttca	35851
gaaatcgcac	taggcaggga	agcctgcctt	accgtcagaa	ttggcagtc	cagtcgaatc	35911
atttccggat	tatcttgtac	attcaatgct	acactagcta	tatccgagtt	atattcgata	35971
gtttgcagggt	tttgtaaaaa	cgacaaaactc	tgtagattag	tgttccgaat	tgcaatagat	36031
cctcgaatca	ttgtgacatt	caaaaatgaa	tcataatcga	aggttgcaat	aatattcact	36091
aaatttagac	cagaatctag	agttttgcat	ttggagtact	ccttaacatt	tgatacatta	36151
actttttcac	catcacatcc	tgaaatttga	ctatttttat	actgttaaaa	aattgtttct	36211
caccacaatc	ctttaagttc	cctctgacaa	tgagctcatt	atacatgtgt	aaaaaqccgc	36271

59/92

FIGURE 19

```
catcacagga aaattccagt ttccgattat tctcgattct aatatcacac gcctcgatac 36331
cccgatcacg gtacaagtag agatcgtaga gcacactggg gtcgtttaat tgtgaattgt 36391
ttccgatgta aacaccgtct gaaatctgaa gtttaagaaa aaattaagta agttttaatc 36451
tacatggtga tccgtttttg ttgaaagtat caaaaaatta actggagtca gaatgtctca 36511
tttcgttttg atcttcaaaa aatgcgggag ttcagaccta gacatctcgt ctgatttcgc 36571
atggttaaga gcgttctgac gtcacaattt ttctgaaaaa atattcccgc attttttgta 36631
gatcaaatta aaatgagaca gcctgacacc acgtggagtt ccttatatac aaaaaagttg 36691
atttttcgct cgtgattttt cgttgtaaca tcatgaaaaa tccagtgttc tctgcaaacc 36751
actaaaatcc acttttttgt ttcagccgct ccgcaagcag cttcgtcgag gtcatggcag 36811
cggccgagtt tcccactccg ctgaaactcg gcacttaata tatgaacgac taagctagca 36871
gggccgccat tctacettac cagcaaaaat gaattcggtc acttacacac atcacacacc 36931
acattaaagt ttcttttttc tttgtcagct gtaaaaaccg aaaggcttgt cagactagta 36991
ttctcaatat taaatc                                     37007
```

FIGURE 20A

ssl-1 Predicted exons:

<u>Exon</u>	<u>Position in genomic sequence (inclusive)</u>
1	1001-1281
2	1923-2027
3	2084-2312
4	4420-5205
5	5855-6487
6	7685-8515
7	9700-10184
8	12211-13165
9	13643-13726
10	13796-13939
11	18879-19101
12	20449-20735
13	21661-22273

Figure 20B

ss1-1 cDNA

```

atgccggcaa caccggtgcg tgcttcaagt actcgaataa gcagacgtac atcatcaaga 60
tcagtggctg atgacagacc atcaacttcg tctgcggtgg ctccacctcc ttcacccatt 120
gccatagaaa ctgatgaaga tgcggtagtt gaggaggaga aaaagaagaa aaagacatca 180
gatgatttgg aaattatcac tccaagaact ccagtcgacg ggcgaattcc ctacatttgc 240
tcgattcttt tgactgaaaa tcgatcgatt cgcgataaat tggttctgag cagcggttcca 300
gttcgtcaag aagatcacga agaacagatt gctcgagctc aacggataca gccagttgtc 360
gatcaaattc aacgagtcga gcaaatcata ctcaatggtt cagtggaaga tattctgaaa 420
gatcctcgat tcgcagtaat ggcagatctc acaaaagaac caccaccaac acctgcacct 480
ctcctccaa tcagaagac aatgcaaccg attgaggtga aaattgagga ttcagagggc 540
tcaaatacgg ctcaaccgag tgttctgccc agttgtggag gaggagagac gaatgtggaa 600
agagccgcca aaagagaagc gcatgtattg gctcgaatcg ccgagctccg taagaacggc 660
ttatggtcga acagtcgtct gccaaagtgc gtcgaacctg aacgtaataa aacgcattgg 720
gattatctac tggaaagagg caaatggatg gcagttgatt tccgaaccga gacgaatagc 780
aagcgaaaaa tcgccaaagt tatagctcac gccattgcga aacagcaccg cgacaagcag 840
atcgagattg agagagccgc cgaacgggag atcaaggaga agcgaaaaaat gtgtgcagga 900
atcgcgaaaga tggtagcgga tttctggtcg tctacggata aagttgtgga tattcgagcg 960
aagggaagttc tggagtcgag gctcaggaag gcgagaaata agcatttgat gtttgtaatt 1020
ggacaagtcg atgaaatgag caatatgtg caagaaggac ttgtttcatc gtcgaaatcc 1080
ccatcaattg catcgatcgc agatgataaa gatgaagaat tcaaagcacc tggctctgat 1140
tcagaatctg acgatgagca gacaattgca aacgcggaaa agtcacagaa aaaggaagat 1200
gttcgacagg aagttgatgc tcttcaaaac gaggcaactg tggatatgga tgactttttg 1260
tacactttac cgccggaata tctgaaggct tatggtctga cgcaggagga tttggaggag 1320
atgaagcgcg agaaattgga ggagcagaag gctcggaaag aagcttgtgg tgataatgag 1380
gagaaaaatg agattgatga aagcccatca tcagatgctc aaaagccttc cacctcaagc 1440
tcagatctca ccgccgagca gcttcaagat ccaacagctg aagacggcaa cggtagtggt 1500
catggtgtac ttgaaaacgt ggattacgtg aagctcaaca gtcaggatag tgatgaacga 1560
caacaagagt tggcgaatat cgcagaagaa gcgctgaaat tccagccaaa aggatataca 1620
cttgagacga cacaagtcaa gacgcccgtc ccattcctga ttcgaggaca actgagagaa 1680
tatcaaatgg ttggattgga ttggatgggt acactttatg agaagaattt gaatggaatt 1740
cttgccgacg agatgggcct gggaaagacg attcaaacga tttccctgct ggctcatatg 1800
gcttgtagtg aatcgatttg gggaccacac gatattgttg tgccgacgtc tgtcattctg 1860
aattgggaga tgggaattcaa gaaatgggtg ccggctctga agattttgac gtattttgtg 1920
acggcggaag agcgtgccga gaagcggaag ggatggatga agccgaattg tttccatgtg 1980
tgcatcacat catacaagac ggttactcaa gatattagag cttttaagca gagggcctgg 2040
cagtacctaa ttctcgatga agctcaaaat atcaaaaact ggaagtccca acgttggcag 2100
gctcttctga atgtccgtgc tcgacgtcgc cttctcctga ccggaactcc acttcagaac 2160
tctctaattg aactgtggtc gttgatgcat tttttgatgc caacaatatt ctcaagtcac 2220
gatgatttca aggattgggt ctcgaatccg ttgacaggga tgatggaagg aaatatggaa 2280
ttcaatgtc cactaatcgg acgacttcac aaagtgtctc gtccgtttat tctgcggcgg 2340
ctcaagaagg aagttgagaa gcagctgccg gagaagactg agcatattgt gaattgttcg 2400
ttgtcaaagc ggcagagata cctgtacgat gactttatga gtcgtagatc acaaaaggag 2460
aatctaaagt ctggaaatat gatgtcgggtg ctcaacattg tgatgcaact ccgaaaatgt 2520
tgtaatcatc cgaatctctt cgagccgagg ccagttgttg ctccgttcgt cgttgagaag 2580
cttcagctcg atgttccggc tcttctcttt gaaatttcgc agcaagatcc ctctcctcc 2640
tcagctagtc aaattccgga aattttcaat ttatccaaaa tcggctatca atcttccgtt 2700
cgatctgcaa aaccactcat cgaagagctt gaagcaatga gcacttatcc ggagccacga 2760
gcaccagaag ttggcggatt tcggttcaat cggacggctt ttgttgcaaa gaatccgcat 2820
acggaagagt cggaggacga aggtgttatg agaagtcgtg ttctgccaaa accaattaat 2880
ggaacagctc aaccacttca aaatggaaat tcaataccac aaaatgctcc aaatcggtcca 2940
caaacttcat gcattcgttc aaaaaccgtc gtaaatcacg ttccactgac catctccacc 3000
gatcgaagtg gttttcattt taatatggcc aatgttgga gagggtgtgt tcgtttggat 3060
gattcagcac gtatgagccc accgctcaaa cgtcagaagc tcaccggaac tgcaacgaat 3120

```


Figure 20B

```

tggagtgatt atgttccgcg acacgttggt gaaaagatgg aagaatcgag aaaaaaccag 3180
ctggaaattg ttcgaaggcg atttgagatg attcgtgctc cgattattcc actggaaatg 3240
gttgcgctgg ttcgagagga aattattgca gaatttccac gtttggtgtt ggaagaggac 3300
gaggttggtc aggagaggct tttggagtat tgcgagttgt tggtgcaaag attcggaatg 3360
tacgtcgaac cagtgtctgac cgatgcttgg cagtgtcgtc catcatcgtc tggctctcca 3420
tcatatattc gcaacaattt atcaaatatc gagctgaatt ctctgtctct tctcctcaac 3480
acctccacta atttcgatag ccgaatgtcg atctcacgtg ctcttcaatt cccagaactc 3540
cgtctgatcg agtacgattg tggaaagctt cagacgttgg ctgttctgct tcgtcagttg 3600
tacctgtaca agcacagatg tctgatcttc acgcaaagtgt caaagatgct cgacgttctg 3660
cagaccttcc tttctcatca cggttatcag tatttccgcc tcgacggtac cactggtgtc 3720
gaacaaagac aggcgagatg ggagcgggtc aacgcggatc ccaagggtgt ttgcttcatt 3780
ctgtcgacga gatccggtgg tgttgagtc aatctaaccg gtgctgacac tgtgatcttc 3840
tacgattcgg attggaatcc gacgatggat gctcaggctc aggatagatg tcatcgtatc 3900
ggacagacga ggaatgtctc gatttatcga ttgatttccg agcgaacaat tgaggagaat 3960
attctgagaa aggcaacaca gaagcggcga cttggagagt tggcaattga cgaggctggc 4020
ttcacacccg agttcttcaa acaatctgac agtattcggg atctttttga tggagagaat 4080
gtggaagtga ctgctgtggc agatgttgcg acgacgatga gcgagaaaga aatggagggt 4140
gcgatggcaa agtgtgaaga tgaagctgat gtgaatgcgg cgaagattgc ggtggccgag 4200
gcgaacgttg ataatgcgga gtttgatgag aatcattgc cgcgatgag caatttgcaa 4260
ggagatgagg aggtgatga gaagtatatg gagttgatac aacagctcaa accaatcgaa 4320
cgatatgcca ttaactttct tgagacacag tacaagccag aatttgagga agaatgcaa 4380
gaggcagagg ctcttatcga caaaaacgc gaagaatggg acaaaaatct caacgatacc 4440
gccgtcattg acctcgacga ttcggatagt ctgctgtca acgatccttc gacttctgcc 4500
gatttttatac agagctcaag tcttttagac gagataaaat tctacgacga gctggacgat 4560
atcatgccaa tctggcttcc accatcacca ccagattcgg atgcggattt cgacttgaga 4620
atggaagatg attgtctcga tctgatgtat gaaattgaac aaatgaacga ggctcgctta 4680
ccacaagttt gtcataaaat gagacgtccg ttggctgaaa aacagcagaa acagaacacg 4740
ttgaatgcgt ttaatgacat tctatcggca aaagaaaagg aatcgggtga cgatgcgggtc 4800
aacaagtgcc ttcaaatgcc acaatccgaa gcgatcacag cagaatctgc agcgtctcca 4860
gcatacacgg aacactcatc attctcgatg gatgatacaa gccaggatgc gaagattgag 4920
ccaagtttga ctgaaaatca acaaccacc accaccgcca ctactactac tacagtacc 4980
caacaacaac aacaacagca gcagcaaaaa tcgtcgaaaa agaagagaaa tgataatcga 5040
acggctcaaa atcgaaacgc tgaaaatggg gtgaaacgag cgacaactcc accaccatca 5100
tggcgtgaag agccagatta tgatggagcc gaattgaaata tagttgaaga ttatgacta 5160
cttcaagcag ttcaagtcga atttgcaaat gctcatttag tcgaaaaatc ggcgaatgag 5220
ggaatggtgt tgaactggga attcgtgtcg aatgccgtta ataagcagac aagatttttc 5280
cgctcggccc gtcaatgctc aattcgatat caaatgtttg ttcggccaaa agagctcgga 5340
cagttggtgg cttctgatcc gatttccaag aaaacgatga aagtcgacct atcgatact 5400
gaattatctc atttgagaaa aggacgaatg actacggaga gccaatatgc tcatgattat 5460
ggaatattga ctgataagaa acatgtgaat agatttaaaa gtgttcgagt ggcggcaaca 5520
cggagacctg ttcagttttg gagaggccct aaaggtagag gaggatgggt tcataatagt 5580
cactgcaact ttttctcac gagggacgag aaaaagtggg ttctaggcca tggccgaggt 5640
gccgacaagt ttcagc
5656

```

63/92

FIGURE 21

ssl-1 protein

<400> 3

```

Met Pro Ala Thr Pro Val Arg Ala Ser Ser Thr Arg Ile Ser Arg Arg
 1           5           10           15
Thr Ser Ser Arg Ser Val Ala Asp Asp Gln Pro Ser Thr Ser Ser Ala
      20           25           30
Val Ala Pro Pro Pro Ser Pro Ile Ala Ile Glu Thr Asp Glu Asp Ala
      35           40           45
Val Val Glu Glu Glu Lys Lys Lys Lys Thr Ser Asp Asp Leu Glu
      50           55           60
Ile Ile Thr Pro Arg Thr Pro Val Asp Arg Arg Ile Pro Tyr Ile Cys
65           70           75           80
Ser Ile Leu Leu Thr Glu Asn Arg Ser Ile Arg Asp Lys Leu Val Leu
      85           90           95
Ser Ser Gly Pro Val Arg Gln Glu Asp His Glu Glu Gln Ile Ala Arg
      100          105          110
Ala Gln Arg Ile Gln Pro Val Val Asp Gln Ile Gln Arg Val Glu Gln
      115          120          125
Ile Ile Leu Asn Gly Ser Val Glu Asp Ile Leu Lys Asp Pro Arg Phe
      130          135          140
Ala Val Met Ala Asp Leu Thr Lys Glu Pro Pro Thr Pro Ala Pro
      145          150          155          160
Pro Pro Pro Ile Gln Lys Thr Met Gln Pro Ile Glu Val Lys Ile Glu
      165          170          175
Asp Ser Glu Gly Ser Asn Thr Ala Gln Pro Ser Val Leu Pro Ser Cys
      180          185          190
Gly Gly Gly Glu Thr Asn Val Glu Arg Ala Ala Lys Arg Glu Ala His
      195          200          205
Val Leu Ala Arg Ile Ala Glu Leu Arg Lys Asn Gly Leu Trp Ser Asn
      210          215          220
Ser Arg Leu Pro Lys Cys Val Glu Pro Glu Arg Asn Lys Thr His Trp
      225          230          235          240
Asp Tyr Leu Leu Glu Val Lys Trp Met Ala Val Asp Phe Arg Thr
      245          250          255
Glu Thr Asn Thr Lys Arg Lys Ile Ala Lys Val Ile Ala His Ala Ile
      260          265          270
Ala Lys Gln His Arg Asp Lys Gln Ile Glu Ile Glu Arg Ala Ala Glu
      275          280          285
Arg Glu Ile Lys Glu Lys Arg Lys Met Cys Ala Gly Ile Ala Lys Met
      290          295          300
Val Arg Asp Phe Trp Ser Ser Thr Asp Lys Val Val Asp Ile Arg Ala
      305          310          315          320
Lys Glu Val Leu Glu Ser Arg Leu Arg Lys Ala Arg Asn Lys His Leu
      325          330          335
Met Phe Val Ile Gly Gln Val Asp Glu Met Ser Asn Ile Val Gln Glu
      340          345          350
Gly Leu Val Ser Ser Ser Lys Ser Pro Ser Ile Ala Ser Asp Arg Asp
      355          360          365
Asp Lys Asp Glu Glu Phe Lys Ala Pro Gly Ser Asp Ser Glu Ser Asp
      370          375          380
Asp Glu Gln Thr Ile Ala Asn Ala Glu Lys Ser Gln Lys Lys Glu Asp
      385          390          395          400
Val Arg Gln Glu Val Asp Ala Leu Gln Asn Glu Ala Thr Val Asp Met
      405          410          415

```

64/92

FIGURE 21

```

Asp Asp Phe Leu Tyr Thr Leu Pro Pro Glu Tyr Leu Lys Ala Tyr Gly
      420      425      430
Leu Thr Gln Glu Asp Leu Glu Glu Met Lys Arg Glu Lys Leu Glu Glu
      435      440      445
Gln Lys Ala Arg Lys Glu Ala Cys Gly Asp Asn Glu Glu Lys Met Glu
      450      455      460
Ile Asp Glu Ser Pro Ser Ser Asp Ala Gln Lys Pro Ser Thr Ser Ser
      465      470      475      480
Ser Asp Leu Thr Ala Glu Gln Leu Gln Asp Pro Thr Ala Glu Asp Gly
      485      490      495
Asn Gly Asp Gly His Gly Val Leu Glu Asn Val Asp Tyr Val Lys Leu
      500      505      510
Asn Ser Gln Asp Ser Asp Glu Arg Gln Gln Glu Leu Ala Asn Ile Ala
      515      520      525
Glu Glu Ala Leu Lys Phe Gln Pro Lys Gly Tyr Thr Leu Glu Thr Thr
      530      535      540
Gln Val Lys Thr Pro Val Pro Phe Leu Ile Arg Gly Gln Leu Arg Glu
      545      550      555      560
Tyr Gln Met Val Gly Leu Asp Trp Met Val Thr Leu Tyr Glu Lys Asn
      565      570      575
Leu Asn Gly Ile Leu Ala Asp Glu Met Gly Leu Gly Lys Thr Ile Gln
      580      585      590
Thr Ile Ser Leu Leu Ala His Met Ala Cys Ser Glu Ser Ile Trp Gly
      595      600      605
Pro His Leu Ile Val Val Pro Thr Ser Val Ile Leu Asn Trp Glu Met
      610      615      620
Glu Phe Lys Lys Trp Cys Pro Ala Leu Lys Ile Leu Thr Tyr Phe Gly
      625      630      635      640
Thr Ala Lys Glu Arg Ala Glu Lys Arg Lys Gly Trp Met Lys Pro Asn
      645      650      655
Cys Phe His Val Cys Ile Thr Ser Tyr Lys Thr Val Thr Gln Asp Ile
      660      665      670
Arg Ala Phe Lys Gln Arg Ala Trp Gln Tyr Leu Ile Leu Asp Glu Ala
      675      680      685
Gln Asn Ile Lys Asn Trp Lys Ser Gln Arg Trp Gln Ala Leu Leu Asn
      690      695      700
Val Arg Ala Arg Arg Arg Leu Leu Leu Thr Gly Thr Pro Leu Gln Asn
      705      710      715      720
Ser Leu Met Glu Leu Trp Ser Leu Met His Phe Leu Met Pro Thr Ile
      725      730      735
Phe Ser Ser His Asp Asp Phe Lys Asp Trp Phe Ser Asn Pro Leu Thr
      740      745      750
Gly Met Met Glu Gly Asn Met Glu Phe Asn Ala Pro Leu Ile Gly Arg
      755      760      765
Leu His Lys Val Leu Arg Pro Phe Ile Leu Arg Arg Leu Lys Lys Glu
      770      775      780
Val Glu Lys Gln Leu Pro Glu Lys Thr Glu His Ile Val Asn Cys Ser
      785      790      795      800
Leu Ser Lys Arg Gln Arg Tyr Leu Tyr Asp Asp Phe Met Ser Arg Arg
      805      810      815
Ser Thr Lys Glu Asn Leu Lys Ser Gly Asn Met Met Ser Val Leu Asn
      820      825      830
Ile Val Met Gln Leu Arg Lys Cys Cys Asn His Pro Asn Leu Phe Glu
      835      840      845
Pro Arg Pro Val Val Ala Pro Phe Val Val Glu Lys Leu Gln Leu Asp
      850      855      860
Val Pro Ala Arg Leu Phe Glu Ile Ser Gln Gln Asp Pro Ser Ser Ser

```

65/92

FIGURE 21

865		870		875		880									
Ser	Ala	Ser	Gln	Ile	Pro	Glu	Ile	Phe	Asn	Leu	Ser	Lys	Ile	Gly	Tyr
		885		890		895									
Gln	Ser	Ser	Val	Arg	Ser	Ala	Lys	Pro	Leu	Ile	Glu	Glu	Leu	Glu	Ala
		900		905		910									
Met	Ser	Thr	Tyr	Pro	Glu	Pro	Arg	Ala	Pro	Glu	Val	Gly	Gly	Phe	Arg
		915		920		925									
Phe	Asn	Arg	Thr	Ala	Phe	Val	Ala	Lys	Asn	Pro	His	Thr	Glu	Glu	Ser
		930		935		940									
Glu	Asp	Glu	Gly	Val	Met	Arg	Ser	Arg	Val	Leu	Pro	Lys	Pro	Ile	Asn
		945		950		955									
Gly	Thr	Ala	Gln	Pro	Leu	Gln	Asn	Gly	Asn	Ser	Ile	Pro	Gln	Asn	Ala
		965		970		975									
Pro	Asn	Arg	Pro	Gln	Thr	Ser	Cys	Ile	Arg	Ser	Lys	Thr	Val	Val	Asn
		980		985		990									
Thr	Val	Pro	Leu	Thr	Ile	Ser	Thr	Asp	Arg	Ser	Gly	Phe	His	Phe	Asn
		995		1000		1005									
Met	Ala	Asn	Val	Gly	Arg	Gly	Val	Val	Arg	Leu	Asp	Asp	Ser	Ala	Arg
		1010		1015		1020									
Met	Ser	Pro	Pro	Leu	Lys	Arg	Gln	Lys	Leu	Thr	Gly	Thr	Ala	Thr	Asn
		1025		1030		1035									
Trp	Ser	Asp	Tyr	Val	Pro	Arg	His	Val	Val	Glu	Lys	Met	Glu	Glu	Ser
		1045		1050		1055									
Arg	Lys	Asn	Gln	Leu	Glu	Ile	Val	Arg	Arg	Arg	Phe	Glu	Met	Ile	Arg
		1060		1065		1070									
Ala	Pro	Ile	Pro	Leu	Glu	Met	Val	Ala	Leu	Val	Arg	Glu	Glu	Glu	Ile
		1075		1080		1085									
Ile	Ala	Glu	Phe	Pro	Arg	Leu	Ala	Val	Glu	Glu	Asp	Glu	Val	Val	Gln
		1090		1095		1100									
Glu	Arg	Leu	Leu	Glu	Tyr	Cys	Glu	Leu	Leu	Val	Gln	Arg	Phe	Gly	Met
		1105		1110		1115									
Tyr	Val	Glu	Pro	Val	Leu	Thr	Asp	Ala	Trp	Gln	Cys	Arg	Pro	Ser	Ser
		1125		1130		1135									
Ser	Gly	Leu	Pro	Ser	Tyr	Ile	Arg	Asn	Asn	Leu	Ser	Asn	Ile	Glu	Leu
		1140		1145		1150									
Asn	Ser	Arg	Ser	Leu	Leu	Leu	Asn	Thr	Ser	Thr	Asn	Phe	Asp	Thr	Arg
		1155		1160		1165									
Met	Ser	Ile	Ser	Arg	Ala	Leu	Gln	Phe	Pro	Glu	Leu	Arg	Leu	Ile	Glu
		1170		1175		1180									
Tyr	Asp	Cys	Gly	Lys	Leu	Gln	Thr	Leu	Ala	Val	Leu	Leu	Arg	Gln	Leu
		1185		1190		1195									
Tyr	Leu	Tyr	Lys	His	Arg	Cys	Leu	Ile	Phe	Thr	Gln	Met	Ser	Lys	Met
		1205		1210		1215									
Leu	Asp	Val	Leu	Gln	Thr	Phe	Leu	Ser	His	His	Gly	Tyr	Gln	Tyr	Phe
		1220		1225		1230									
Arg	Leu	Asp	Gly	Thr	Thr	Gly	Val	Glu	Gln	Arg	Gln	Ala	Met	Met	Glu
		1235		1240		1245									
Arg	Phe	Asn	Ala	Asp	Pro	Lys	Val	Phe	Cys	Phe	Ile	Leu	Ser	Thr	Arg
		1250		1255		1260									
Ser	Gly	Gly	Val	Gly	Val	Asn	Leu	Thr	Gly	Ala	Asp	Thr	Val	Ile	Phe
		1265		1270		1275									
Tyr	Asp	Ser	Asp	Trp	Asn	Pro	Thr	Met	Asp	Ala	Gln	Ala	Gln	Asp	Arg
		1285		1290		1295									
Cys	His	Arg	Ile	Gly	Gln	Thr	Arg	Asn	Val	Ser	Ile	Tyr	Arg	Leu	Ile
		1300		1305		1310									
Ser	Glu	Arg	Thr	Ile	Glu	Glu	Asn	Ile	Leu	Arg	Lys	Ala	Thr	Gln	Lys
		1315		1320		1325									

66/92

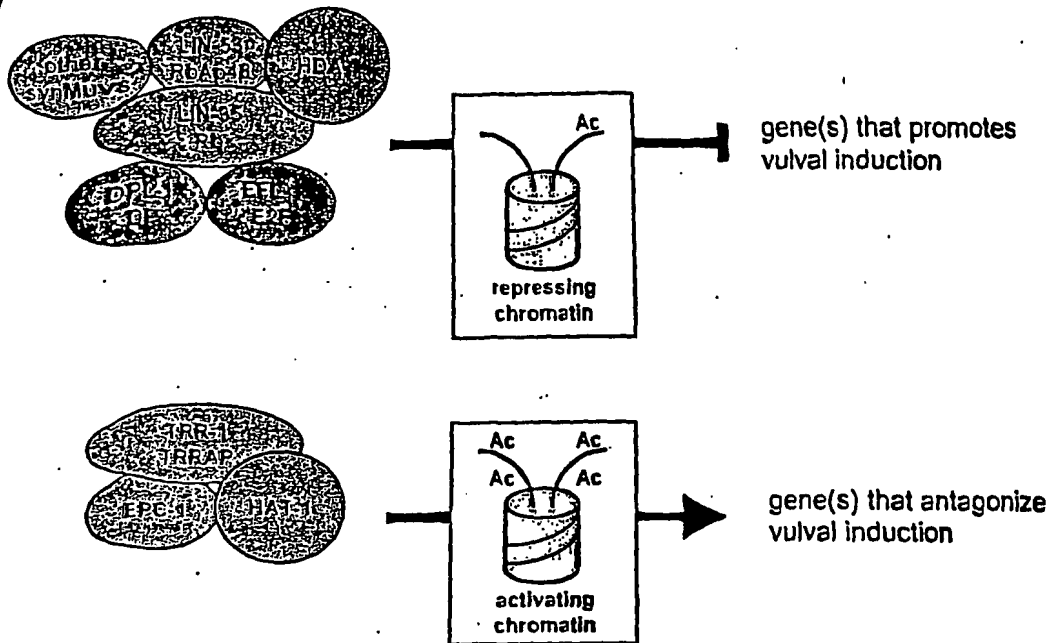
FIGURE 21

Arg Arg Leu Gly Glu Leu Ala Ile Asp Glu Ala Gly Phe Thr Pro Glu
 1330 1335 1340
 Phe Phe Lys Gln Ser Asp Ser Ile Arg Asp Leu Phe Asp Gly Glu Asn
 1345 1350 1355 1360
 Val Glu Val Thr Ala Val Ala Asp Val Ala Thr Thr Met Ser Glu Lys
 1365 1370 1375
 Glu Met Glu Val Ala Met Ala Lys Cys Glu Asp Glu Ala Asp Val Asn
 1380 1385 1390
 Ala Ala Lys Ile Ala Val Ala Glu Ala Asn Val Asp Asn Ala Glu Phe
 1395 1400 1405
 Asp Glu Lys Ser Leu Pro Pro Met Ser Asn Leu Gln Gly Asp Glu Glu
 1410 1415 1420
 Ala Asp Glu Lys Tyr Met Glu Leu Ile Gln Gln Leu Lys Pro Ile Glu
 1425 1430 1435 1440
 Arg Tyr Ala Ile Asn Phe Leu Glu Thr Gln Tyr Lys Pro Glu Phe Glu
 1445 1450 1455
 Glu Glu Cys Lys Glu Ala Glu Ala Leu Ile Asp Gln Lys Arg Glu Glu
 1460 1465 1470
 Trp Asp Lys Asn Leu Asn Asp Thr Ala Val Ile Asp Leu Asp Asp Ser
 1475 1480 1485
 Asp Ser Leu Leu Leu Asn Asp Pro Ser Thr Ser Ala Asp Phe Tyr Gln
 1490 1495 1500
 Ser Ser Ser Leu Leu Asp Glu Ile Lys Phe Tyr Asp Glu Leu Asp Asp
 1505 1510 1515 1520
 Ile Met Pro Ile Trp Leu Pro Pro Ser Pro Pro Asp Ser Asp Ala Asp
 1525 1530 1535
 Phe Asp Leu Arg Met Glu Asp Asp Cys Leu Asp Leu Met Tyr Glu Ile
 1540 1545 1550
 Glu Gln Met Asn Glu Ala Arg Leu Pro Gln Val Cys His Glu Met Arg
 1555 1560 1565
 Arg Pro Leu Ala Glu Lys Gln Lys Gln Asn Thr Leu Asn Ala Phe
 1570 1575 1580
 Asn Asp Ile Leu Ser Ala Lys Glu Lys Glu Ser Val Tyr Asp Ala Val
 1585 1590 1595 1600
 Asn Lys Cys Leu Gln Met Pro Gln Ser Glu Ala Ile Thr Ala Glu Ser
 1605 1610 1615
 Ala Ala Ser Pro Ala Tyr Thr Glu His Ser Ser Phe Ser Met Asp Asp
 1620 1625 1630
 Thr Ser Gln Asp Ala Lys Ile Glu Pro Ser Leu Thr Glu Asn Gln Gln
 1635 1640 1645
 Pro Thr Thr Thr Ala Thr Thr Thr Thr Val Pro Gln Gln Gln Gln
 1650 1655 1660
 Gln Gln Gln Gln Gln Lys Ser Ser Lys Lys Lys Arg Asn Asp Asn Arg
 1665 1670 1675 1680
 Thr Ala Gln Asn Arg Thr Ala Glu Asn Gly Val Lys Arg Ala Thr Thr
 1685 1690 1695
 Pro Pro Pro Ser Trp Arg Glu Glu Pro Asp Tyr Asp Gly Ala Glu Trp
 1700 1705 1710
 Asn Ile Val Glu Asp Tyr Ala Leu Leu Gln Ala Val Gln Val Glu Phe
 1715 1720 1725
 Ala Asn Ala His Leu Val Glu Lys Ser Ala Asn Glu Gly Met Val Leu
 1730 1735 1740
 Asn Trp Glu Phe Val Ser Asn Ala Val Asn Lys Gln Thr Arg Phe Phe
 1745 1750 1755 1760
 Arg Ser Ala Arg Gln Cys Ser Ile Arg Tyr Gln Met Phe Val Arg Pro
 1765 1770 1775
 Lys Glu Leu Gly Gln Leu Val Ala Ser Asp Pro Ile Ser Lys Lys Thr

5 of 5

FIGURE 22

A)



B)

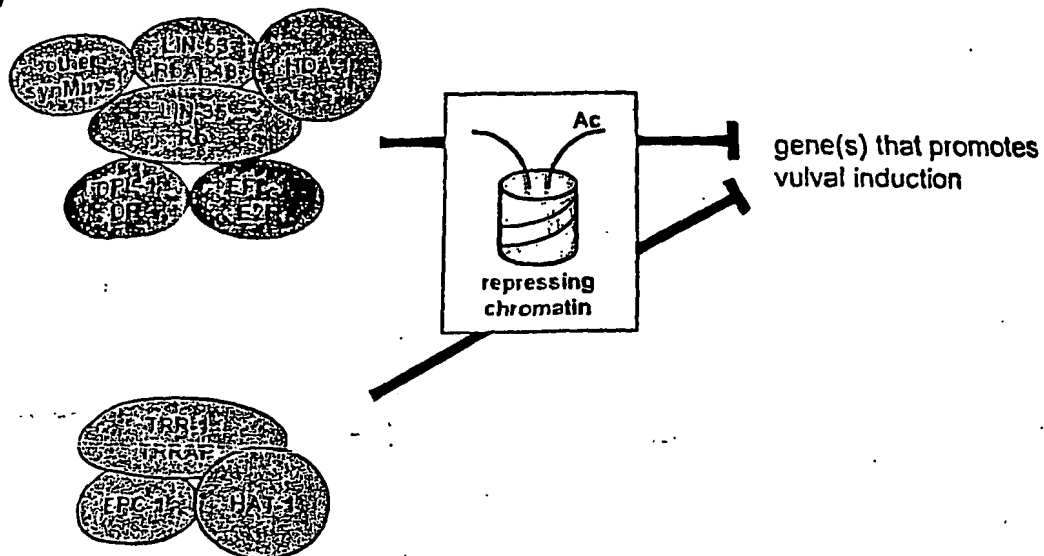


Figure 23

lin(n3628) genomic sequence (1 kb of upstream and downstream genomic sequence is included in this file).

<u>Exon number</u>	<u>Exon boundaries (inclusive)</u>
1	1001 – 1035
2	1920 – 2062
3	2114 – 2190
4	2241 – 2501
5	2551 – 2903
6	2955 – 3405
7	3497 – 3631
8	4227 – 4690
9	5293 – 6058
10	6696 – 7058
11	7609 – 8338
12	8771 – 8933
13	9511 – 10306
14	10774 – 10851

GTCAATGGAATTCTCGACGCGGATCTTGTTAGAGATGCCGTCGAGAGAGATT
TGATCAAATTGCGGTACGCTGAAACGGATGCACCAGTTTTACAGGTAAAATG
GAAATATACAACTCAAAAGTAAAATTTTATGAATTTTCAGATCAACAACTCA
CTATACACGGCATCCTGGGAGCAAGATCTCGGAACAAATATGGTTCTGCAGT
CAAAAGGAAAAGAGATGGAAGTGATTTCTGTGTACATCGACCATGATGACTGC
AGAAAAAGCCCTGTTGACCTCGTTAAGCACCGAAGGATCTACACTAGCCGCC
AATGCAGAGACTGCTCCGAAATCTGATCTCAGTCGAACTCAACCACGTCAAC
AATGATTTTCAAAATATAAATTAACATGAAGCTCTGAAATAAACTCATATAA
CTGCTAAAATAAACTGTTGCTTTTGAAACCAACATTTGTTAGACAACCTGCG
TCTCACAGTCATTTTTCAATATATTGGCGCCGCGCACACACAAAGAAGAAGA
ATTCGTCTCATGGCATGGCATGTGCAGTCAGCGGCCACCCTGTGTAACCACT
GCGTATCGCATCTTTCCACGTGTTTTTGCAATCTTGCTGTCACGTTCAATTCCT
CGTACAACCATCTCTTCTACCCCGTTGCCTCCTCCACCATCTCATCTCAATTG
TGTCGTTGCCCTCCCTCTCCCAAGTCTTCTGCGTCTCTTAGTGCTCTTCGAG
AAAAGAACGAGGAGAGCTGTGAGACGCTAGTAGGAAACGCATTCTCAATTC
GATATAGGCACATTGAGAGAGAGCGAGCGCGTTTCGACGTCTTCTAGCCTT
CACATCATCCAGACGACGTTACACGCACACACAGCCAACCCACCCTTCTG
ACAACGAATAGACGACGAAGAAGAGAAGAAGAAAAAGAAGAAGGTACCCA
TTTTTCATTCCCTTTTTGCCTCCACACTTCACTATTATCGATTTTGTGAGCGAG
CTCTAATGTTTCAACGCAAAGTGGTATTGCCTAAAAAGCGGTGAGAATTTGCT
TCAGACAGAAATTCGTTTTTTTAAACAAGAAAAATCCGGTTTCAATTGTCGTA
GAAGGTCAATTTTACTTTCAACGCTCTTCATTGACGGAAAACGTTTTTCTT
TCAAATTTTAAATTACAGAGGCATTTTACTCAAGGTTTGTTTTAATTTAAATT
AAAAATAAATTTTAAATAGAAATATGGATAATATAAAATGTTTTCTTCAA
AAATGCACTCAGGTTACCAAAAAATCGATAATTAAAAATACGGTCGCAAAG
GAGCGTCGTTAGCTGCTAATCAATGGTCTTAAAACGAATCTATCGATTTTG
TGTAACACACGACGAAGTGCTCCACCGTTATTTTTTGAACGAGTGCGTTGC

FIGURE 23

AATTCCATCCCATTTTGACGTTTTCTTTTTTTTTTCATCAAATTTTTTAGCATT
TAAAGTAAAGTCAATGATAACCTGCAAATAATAATGTAAAATTCATTAATAAA
CCGAGAGAAAAAGTCTAAAGTCATAAATTTTTGATAAAAAAGTGATTTTCGA
AACTAAAAATCATTCAAATTAAGTTGAACCTGATTCTTCAATTTTTATTATA
TATTAAGGCTTGATCCACTCAAATAAAAGGAGTTTTTAATTGAGAAAAAAA
GCAAATGAAAAAATCGATAATTAATTTGGGCGCCAACCTAGATTTTAATATG
TTTTTGTTAGAAATTTGTATATTTTCATCACTCTCTGACTTTAAGCATTCTGTAT
TTTAAGGAAGTGTGAGCTTTCTAATATGTTTTTTTATTAAAAAAACATGTTTT
TAACAATCTCCCTGTCATCCCCATCACCTAATGCACTCAAATAATCAATAATC
ACAATACTTTTATTTTTTCTTGCGAAGACAGAAATGGTCCAAACGAGACGAAA
GACAGCTGCAGCTGTACAGGACGGTGGTGCCGTTAAGGAGAACAAAGCCAA
GCCACCTGCCCTCAAACGCCTACAAAACGAGCAAAACGAGGTCTGTCCTCCCG
AAAATTAAGACTGGTGAGCGAATGACTATACGGAAGATTGAAAATTCACGTG
GAATACTTGCGAGATGCCAATACTTTGAATACGCCAAGCACTTCTTCCAACCTG
GTCGATGACAACTTCTCATTGAGTCTGAATCACAGGTAAATTGATTCTTTTC
TATTCAAAAATTAATCTAAACTATACATTCCAGGACTCGATTCTCACAAACGA
AGCCGACTCTTTTCTGGAAAAAGAAGTGGAAGAAATCGAAGATAGTTCAGAT
ATACTTCCCGATAAAATTAATTCTCCAGAAAAACCAAGTGTTTTGGTGAAGC
GGAGATCGAGTACGCGGTTAAAAGTGAAGACTGATGAAGATGAAAAAGATG
TTCCTGTGAACATAGAAGTAGCCGTTTTAGAAGAAAAATCAATTCAAATCGA
GCCAACATCTCCCGCTCACCCGGAAGATCCTCAGGTGAGCTTTTTTTAAAAAT
ATGTATTAATCAAAATTCCTTCATTTCCAGCCTTCGACTTCTTCTCTTCCACTG
GTAGAACCAATTGAAGACATTGTGGAGCCAAATGAGCCAACAAGCTCTGCCG
ATCCTCCAGTATCAAATATTAAGGATGAGGATATTAAGAAGAAGAGCCACT
GATTA AAAAGCCAGCTTCCGATGAGTCAGAATCTATGGATATAGCTAACTCT
GAAAGTGGAAATGATTCCGATTCAAGTGAAGCTGATCCTAGGACGATACCAT
CTTTCTCTATACCTCTTCCCGACACACCACCTCCAAATTTTGCGAAAAGAGGA
GAAATACATGTAGATGTAGATCAGAAAAATTCCAAGCAATCAGGAGAATCAC
AATCGCCTTGGGAGCGGTAAGAATATTTATCCTAGCCAGGTGTTATAACAAA
ATTGAATAGTTTCAGAGCAAGAGAAAAAGTCTGCATCGAACCCATTGTCCTCT
CCAACAATGAGCCGACCCAGGATACACTTCCTTCATCCAGCATATCAAAGTTT
CACAAATGATTGAGTTTCACCTCTACCACCACCGCCACCAGAGCCGGCTCCA
GCTCGTGAAAAAGTGGAAAAATGGTGGTCCAACCTACTTTCAAATGACTTTCA
AAAAAGCTGCAAATATTCCTATCTTGAAGACATCGGCATTTGAACAACCATC
ATCACCTCCACCTTCCTCATCAGTTTCTTCATCAATTTCAATTATCTGAAGTGAA
TTCTTCTACATCGATAGCCTCCGAGTCTTCTCCAGCGAAAAGAAGCTCAAATT
TCGATTTAACTGCCTCAAATGAGCTTCCACCACCTCAGATGGTTGAACTTCCC
AAGCTCTCATTTTTCAATATGCCTCCAGCCGTTCCGCTCCGCAGAGGTTAGTTA
ACTTTTTCCCGGTTTCATGAAATTTAGCGGTATCTGTCCTCCTTTTGGTGTGT
GCCCTCACAACTAACCTCTTTTATCCAGGACGATTCTGCGATGACGTCGGAA
GAACCGATCCTTCTCCTCCGTTCTCCGAATTCCGCCACTCCTGATGATGATGC
ACTTTTCTCAGACCCCAACCAACCAAGATGACCGAATCAGAAATTCAA
GCACTGGTGAGCCAGATCACACATTTGATGTCGTGTGTGGAACCCAGGAAT
TTCAGACCGTTTTTCTTTACACCTCATCCCTTTTGTGTTATGTTAACATTTCAT
TTTGTGTCTCAAACACTGCATGCTTTTGCACCTTGGAATTAATAATAATGCG
TTCTGGGATTTTGTGTGTTAAGGTGGAGTAGAGTTTGTGAGGCTAGAAAGTAT
GCCTTTTTCGTTTCTCCACTGCAAAATTTGTTTTGAAAAAACAAAAAATTA
CTAAAATTTGAAATTTACCAACTTGCCGTTGTCACAGCTGCTGAAATACAGT

71/92

FIGURE 23

TTTTATTGCATTTTACCCTTTATTGCATATTATTATTAGACACCTTTTAGGTC
AATAGGCAACCGAAAATATCCGAATTTGACTTAAAATGTACCTAAATTAAGG
AACTAACTTGAGATATACGACTAAAAATGCAATAAATTGTGAGAATTATTGT
TATGAAATTCAGCCGTTTTAGGCTAGTTTTAGCCAAAAACCGACAACTCTAT
TCCAATTAATTTTCCACTCCTGCACCTCGATTAGTGATTTTTTTGAAGAAAAAA
AATTATCTTCTTATTTTCAGAAAGTAGCGACGGAAAAAGTGAATCAAGTAATT
GCTCGACGTGAAGATTCTGAAAAAGATGTACGTCACAGAGAAGATCGAGATG
ATTATGATAGACGACGTGACGACCGTGACAGAAGATCCAGAAAGACTGATTC
GGAACGAAATGATCAAAGAGGACGACAACGTGAAGATGATGAACGAAGAGC
TCGAGAACGAGAAAGAGAAGTTACGAAACGACATGATCGGGAAAGGGAAGA
GATGCGATTACAGAAACAAAAAGATGAGGAAAGAAGAAAGAAAGATGAAG
AGGAAAGGATACAAAAAGAGAATGATGAGAAAAAACAAAAAGAGGATGAA
GCCAAAATGGAGGAGGAGAAAAAGAAGATTAAAGAGGAGGAAATGAAGAT
TCCTGAATTTGAGTTGATTAGCGAATCAAATATTTGACGAGGAATGCGAAT
AAAAAGAAGACTGAATCCTTAACGTAAGTTATTATTTATAAATTTGACTTAAA
AATTGATAACTTTCAAATTAAGTGATTCAATAGACTCAAAGAATGAAAAA
CTAGAGTGCGCCTTTAAAGAGTACTGTAATTTCAAACCTTTTGTTGCTGCTCAT
TTTTCATCGATTTTTCTTAGTTTTTCGTTAAAAATAATTCAACCATTGGATTAA
AAAAAATTAAAAACACATAAATTTTATTTTGAAAAGTAATGAGAAAAACTAT
AGAAATTCGCCGAAAATTCTACAGCAACAAAAGCTCAAATTTACAGTACTTT
TTAAAGGAGCACATCTTTCTGAATTTAACAAAAATTCGGAGATTTTTCTTTTT
TTCGTGTTTTTCTGGCGAAAAAACGATTTTTTCGCTTTTACCGGAAACGGTATC
CGGAGGAAAAAAAACGAAAAAAGCGAAAAATTTTAAGAAGTTTCAAGAT
TAGTTACAACTCTTTTCAAAGCAGATTCTACAGTTTTTTTGGGGTTTTGCCA
AAAAATTTATGAAATATAATGTTTTTTAGACTAGAAAAATAAACTAATTTTAA
TTTTCAATCAAAGCTCATTATTATATTTATATTTATATAATTCAGTTGCGAAT
GCCATCGAACTGGTGGAACTGTTCCGGACAATACTTGTGTGAATCGTGCAAT
GCTCACCGAGTGCCCATCATCATGTCAGGTCAAATGCAAGAATCAACGATTT
GCAAAGAAAAAGTACGCGGCTGTTGAAGCATTCCACACTGGAACCGCCAAA
GGATGTGGACTTCGAGCAGTGAAAGACATAAAAAAAGGAAGATTCATCATTG
AATATATAGGAGAAGTTGTGGAAAGAGATGATTATGAGAAGAGAAAAACGA
AATATGCAGCTGATAAAAAGCACAAACATCATTATCTCTGTGATACTGGAGT
CTACACGATCGACGCAACAGTCTACGGAAATCCATCTCGATTTGTGAATCAT
AGTTGTGATCCTAATGCTATATGTGAGAAATGGTCTGTACCAAGAACTCCTGG
AGACGTTAATCGAGTTGGTTTTCTTCTCGAAACGATTCATTAAAGCCGGCGAA
GAAATCACATTTGATTATCAATTTGTCAACTACGGACGTGACGCTCAACAATG
TTTCTGTGGAAGTGCTTCATGTAGTGATGGATTGGGCAGAAACCGGAAGAA
TTTTCATCTGATGAGGATGATGATATTGTGACTACAAGGCATATTAATATGGA
TGAAGAAGAAGAAGAAAAGTTGGAAGGTCTTGATCATCTTGGAATCATGAA
CGGAATGAAGTGATCAAGGATATGTTGGATGATTTGGTCATTTCGGAATAAGA
AGCATGCTAGGAAGGTTATCACAAATTGCGGTAAGCATTATTTGTAGAGAAA
ATTTAAAAATTAAAGATGGAGTACCGAAATCEGAGAAATATATTTAATTGAC
TCCAATTTTTCTCTGATTCGGAATTTTTAAATGAAAAAATTCAAAAAATTT
CCTTGATTTTATGTTTTAACTTGAAATTGCGAATTTCAATTTGTACAGATTTTTG
AAACGCCGAATTTTCGCGCCAGAGAAGCCATGTGTCGATTTTTGAGATTTGTG
TATATTTACAAGATTTTGAATCTTCATCGGATGCTGATTTGCGTTTTTTCATCAT
TATATTATCAAAAAACTAACAATTTGTTCCGTTTTTACGGAAATTAACAATATA
GACTAGACATTTCTGTAATATACACAAATCTCGTAAATCGACACATGGCGTC

FIGURE 23

TCTGGCGCGAAAAATTCGGCATTGAAAAATCTTATGCGGGCACTAATGAAAT
TCGTGATTTCAAGCTGAAATATAAAATCAGGGAATTTTCCTTGCATTTTTTCA
CTCAGAACTTCGGAATCAGTTGCAAATTTGGAGTCATTTGAAAATATTTCTCA
GATTTCTGGTACTCCACCTTTATTATAATTTTTTAAAATTTTTTAAATGATTTTTT
TTCCATGTTCAACAAAAAATAAAATTTTCAGTCTGCAATGACCGATTACTCTC
AACGTGTGGATGTCATTCAAGAAATCTTCTCCTCAGACACCTCCGTAAACCGTT
CAAAAATTCTATGCAAAAGAGGGAATGGCTACATTGATGGCTGAATGGTTGT
CTGAAGATGATTATTCGCTGGATAATCTGAAACTTGTTCAAGCTATTCTCAA
GCTCTTCACACTGAACTATTCGATTTCGTGCGCCAAAAATGATCGACTCTTACG
AGATTCTACATCAGATGGGTCAATGCGAAAAATGGATGAATATGTTGATATA
CAAGTGATAGCTGATTCACCTATTGCTTGTGTTGAAGATCCCGTACAGGAGTA
CAAGGATGTTTGCAAAGTTATAGAGGTATATACATATTAATTTTTTAAAAAAG
AATATTTTTTGCATGTCACAAAATATTTGGAAATTTTCCCGAAAAACCCATGA
AATCAAAAAACAAATTAAGTAAAAATTAATTCCTCCTACGAACATTTTTCG
ATTTTTTCGTTTTCCGATATTCCTTTTTAAAAATCTGATTTAAAAAATAAACT
TAAATTTTAGGTCTTTTTGCTCCTTTTTAGAAGCAATTTATATGTTTTTTAAAA
CAAACTTAAAATTAGCATTTTTATGGGTAATTTTCTGAACACATTTTTTTTTCT
GAAAAAATGGCCAGAATTTCAACCACTTCTCCGTAAAATCGAAATTAATA
ATTTTTTCTCTATACATTTTTCAAAAAAAGACTCCTCATTTATTGTATTAGATA
CAAATATATGTTTTCTCATCAAAATTTACGAAATTTGTTATAATTTGAATTT
TTTTTGTTTTTTTTTCGAAAAATTGAAAATTTTCTAATTTTGAAACGATATTAT
ACAATTTTCAGCGCCATCAATTTAACTAATTAATAATTTTCAGAAAGGTCTCGT
CGAAAACCTTCACAAGAGCCAAAGAGATGGCCTATCGGTTAAATCAATACTGG
TTCAATCGATCAGTGAGCTTCAAATTTCCAAAAAAGATACGTGATCCTGTGC
CAAAAGATGTTCCAGTCAGACAAGAAGATGCTACAACATCATCACAACTCTCA
TGATAATAGTAGTAGAACTGTATCACCGAATCATCGACATCATTCATCTTCAT
ATTCAAATTCATGTTATCAAGAACGAGAACCATCTCATATACGATTCTTTAAT
AATGGAAATGATGTTTCATCAATATCGTTTTTGGAGGTTATCATGGAAATAACTA
CAATGATAACTATTTTCAGTAGAAGGCCCAATAAGGATTCATATCGAGATCGC
CGTCGATTTAATGGACGTCGTTTCGAGAAGTCGATCAAGAAGTGTCTCACCAC
AGAACTATAAAAGAAGAAAACCTCGATGAACATGACAATAATCATCGTCAGC
GTTCTCCAATTCGTGATCGTCACACATCTCCCGGCGGCGAAAAGACTCCTAGC
TCGAATAATTCTGGAGAACGAACTATAAAAGACTGGATATTCGAGGAGCTC
GTATAAAACTATAAAAGAAGATTTGGAAGCTGCTGCTGCTGCTGCTGCTGC
TGCTGCTGTACCATCAGAAGTGCAAGCTTATCCTCATGAACATACAGCTGTAC
ATCAGAGTGTTTATCAGATGCCAGGTTATGAGTCTTATGGTTGGTTTAGTTTT
TTTTAAAAATATCATTTACCAGGGTGCCATTTTTTAAAAATAAAAAATAACTCGGA
AAATATGTTTTTAAAAAATTTTCAGAAATTTCTCTCATCAACATAAAACTTGATA
AAAATCGAATTTTTATTATTTTCTAAACATTTTTTTCGGTTTTTCCGAAAATCAA
AAAAAAGTTTAGAAAAATAGCAAAAAATCAGTTTATTAGAAATCAAATTTTG
TTCGTTTTTGATAAGAAAAAACATAAGAAAACATGTTATTTTCTTCTGAAAAAA
GAAAAAATCGAAAAATCTATGGCCTTTTGGCAAAATGTTTTGGACCAAAAA
ACAAAACAAATAGCATTAATAATTATTAGTTCTTTTGTTTTCTTCTAAAGTTAA
TTTTCTGAAAGTCTTGCTTGTCGTATATCAAATAAAAAACATTTTTTCAGGAGTA
TATGATCCTGTAAATGGTGTCTACATGTATCCTCATCCTGGCGCTGGTTACTA
TCCACCTGCCTATCCACAACAACCGATTATGTTAACAATGGACACTCTCCAC
CGAATGATCGTCTTGGTGAACCTTACGAGAAAGCCAGTATCGAGCAGCTAGC
GTGAGCATTTTTTAGTTTTAAACCTTTCGGATTTACCTAGAAAAATGTTACCTTT

FIGURE 23

GACGCAAAATTACGGTAGCAGGTCTCGTCGCGACCGAAATTTTTCAGCGGAG
TACGGTAGCTTCCCATGAATTTTTTTGCTGAACTTATCTTTCTGATAACAAATA
GTAATAAAACATGAAAACTGAATAAAAATTGATATCTTTACCTTATAGGC
TCTTTAAGGGCGCAGACACAAAACTGACCGGCTACCGTAATTTTTCGTCAA
AAGTCACACATTTCTCAACTGGTGAAATCCGAAAAAATTGAAATTTTTACTAC
TCGTCCGACTGTTTAGAAAAAGATTAAAAAAGAAAAAAGAATGTCGGTT
TTTCGAATTTTCGATTTTCAAAGAAAAAATCAATATTTAAAAATCATTTCG
GTAATTTCCCTAAATTTGTAAATATAATTTCCAATAAATGTTTTTTGTTTTCC
GGAATTTTAATAAAAAATCAATTTTCGCGTAACAAAAATGCGAAAAAATGAC
TAGCCACTCGAATATAATAACACATGAAATAAAATTAAAAATTATTACAGTCA
ACGAGATGCAATTGTGAGACAAGAAGCTTGAGCTGATACGTATTCAAATCGAA
AGAAAACTGCTCAAAAAGAAGCGATCAAGGCCGCTTGCCGTCGTGCTAACG
AAGAAGAAGCTAAACGACAAGAGGCACTTGCAAAGACGAAATATGTTTGGG
CGATTGCAAAGTCAGAAGCTGGAGAGACGTATTACTACAACAAAATAACAA
AAGAGACGCAGTGGACAGCACCAACACCAAGTTCAAGGTCTTCTCGAACCGGC
TTGTGGTGCATCTCCTGATACTACAGTTGTCAATTGCTGACGAGATTACTGAAG
AAGAGCAACAAGCTGAAGTTCTGGAGAAGCCGCGTGTGTTAAGGAAGAAG
TTATCGAGCCAGGTTCACAATCTGAACTCAAAAAGAATCTCCGGAGAAAGT
TCGAGTTGTTGTACCGAAAGTTGAAGTTGAAAGATCACCGTCGCCAAAATCT
TCTCGTGATCGTGAGAAGGATCGAGAGAAATCTCGTGAGAAAGATCGTGAAA
GAGATCGTGACAGAAGAGAAGGTTCAAAACATCGTGATAGTTATCATGGACA
TCGAAACGGCAGCAGTTCTGTCACTGAACGACGTATGCGAGAGTTCAAACAT
GAGCTGGAACGATCCACTCGATCTGCCGTTCTGTTCTCGTCTACAACATCAACG
TGACGCTTCTAGTGATAAGACTACTTGGCTTATTAAGTTAATATATCGAGAGA
TTTTCAAACGAGAAAGTGCGCAGAGTGGATTTGATTATCGATTCAAGTGAGAA
TACTGATAAGAAGGTAATATTATGGACCAAAAAATAAACAATTGAAAAAAA
AACCAAAAAAATCTGATGCTTGAATTTAAAAAACAATGAAAGAGTGCA
ATTTTTTAGGTTTTTTTGGTCTTTTTTTTTTGGA AAAACCAAAAAATAAATTTTTT
TCCAAAGTACCAAACTTCATTTTAAAAAATTTTATTTGACATAAAAATTGATA
ATTTAAACTAATTTGAACATTTTTCCGCAAAAATTATAGATTTTTCTGCCAA
TTTTAGATTTTTAACGTTTTTTTTTCGGACAATTAATGTTTCGAATCATCAATCA
GAATGAATATGATATCTGATGAAATTCAAAAATAATGCAATTTAAATAGAAA
ACGGTACAAAAGTTTTGAAAAATTTAGAAGAATTCTAAAAAATCCTGTCC
TTCAGGACAAAATTCAACCTTTTTCTCAAAACACAAAAATTACTTTATATTAT
TTTTCAGGTGAAAACTACGTCAAGTCATATATCGACCGAAAACCTCGAATCA
AACGATCTCTGGAAAGAATACTCTCGGCCATGAGCTTTATTTTTTAATTTAAA
TTTTATAAAAAAATGTTTATGCTTGTTTTTTTTCTCTATAGTTCCCTCCTATCCC
CCCCCTCCCCTATCGCCTAAAAATTGATCTCTGTCTGATTTACCGATTTCCGT
TTTATTTGATCCCATTTGAACGAGTATATCATCATGTTTCTGAACCTCAACGTTT
GCACATTTTATCCCCTAGTTTTATGTCCCCAGAATTGTTTTATACTATCCTGT
AATCCACCTCAAAATGACAGCCATGAAAAGCTGTTTTTCATGTTTTCTATTTT
CTTGTTGATCGTATTTGCGCGCGCTCTTTGTCGCGCAAAATTTTTTTTGTAATTA
AAAATGAATTACGGATGTTGAATTTTTAAATTTATTTTTTTTAAAGAAAAATTG
TGGAAGTTTTTTCAGATTCTATACTGCTTATTTTTACGCTAAATTTTTTTTCGAA
GTCCCCTTTTTTCAAATCGAAGTGTAAGTTCGCTCCACGATCAATAGAGACTC
TCCGCCCTCGAACCATGGGTCTCGTTAGGTATTTGGCAGACTTACCGTAAATT
CAAATGTTTTATTACTTCGCGACTAATTTTTTTTATTCATGACTCAATTTTTTAT
CAATCCAACGAAAACTAATTA AAAACAACGGAAAAACATAACGAAAAATG

FIGURE 23

CTTGAAAATTGCAGACATTTCCGAAATTAATTAAATTCCTAACGAGACCCATG
GCTCGGGGGCGGAGTGTTTTTCGATTAGCCATGGAGCGCGTTGAGATATTCCT
AAATTTTTCTATTCAGATGTCGAATCAATCAAAACGGGTCACAGTGAGAATT
GAGCATTCGAAGAACAACCTTTTTTCGAAAAGTAATTTTCAAATTTTGATCCAAA
GAAATTATTCGTCAATTTTCAGAGTTTTAAAATTCCAACATCAAGAGCAAGA
AGATCGGAAGCTCAAATATGTTCTGCACAAAGCTCACGAGAATCTGAGAAAG
TGCCCATTCGAGATTCTGACAATTG

Figure 24 LIN(n3628) Protein

MFQRKVVLPKKRTEMVQTRRKTA AAVQDGGAVKENKAKPPAPQTPTKRAKRG
RPPKIKTDANTLNTPTSTSSNLVDDKLLIESESQDSILTNEADSFLEKEVEEIEDSSDI
LPDKINSPEKPSVLVKRRSSTRLKVKTDEDEKDVVNIEVAVLEEKSIQIEPTSPA
PEDPQPSTSSLPLVEPIEDIVEPNEPTSSADPPVSNIKDEDIKEEPLIKKPASDESES
MDIANSESGNDSDSSEADPRTIPSFSLPDTPPPNAFAKRGEIHVDVDQKNSKQSGE
SQSPWERAREKSASNPLSSPTMSRPRIHFLHPAYQSFTNDSVSPPLPPPPPEPAPARE
KVENGGPSTTFKMTFKKAANPILKTSAFEQPSSPPSSSVSSISLSEVNSSTSIASES
SPAKRSSNFDLTASNELPPPQMVELPKLSFFNMPPAVRSAEDDSAMTSEEPILLR
SPNSATPDDDALFLTTPPPPKMTESEIQALKVATEKVNQVIARREDSEKDVRHRE
DRDDYDRRRDRDRRRSRKTDSENRDQGRQREDDERRAREREREVTKRHDRER
EEMRLQKQKDEERRKKDEERIQKENDEKKQKEDEAKMEEKKKKIKEEEMKIPE
FELISESKYLTRNANKKKTESLTCECHRTGGNCSDNTCVNRAMLTECPSSCQVKC
KNQRFKAKKYYAAVEAFHTGTAKGCGLRVAVKDIKKGRFIEYIGEVVERDDYEKR
KTKYAADKKHKKHHYLCDTGVYTIDATVYGNPSRFVNHSCDPNAICEKWSVPRT
PGDVNRVGGFFSKRFIKAGEEITFDYQFVNYGRDAQQCFCGSASCSGWIGQKPEEF
SSDEDDDIVTTRHINMDEEEEEKLEGLDHLGNHERNEVIKDMMLDDLVRNKKHA
RKVITIASAMTDYSQRVDVIEIFSSDTSVTVQKFYAKEGMATLMAEWLSEDDY
SLDNLKLVQAILKALHTELFDSKAKNDRLLRDSTSRWVNAKMDEYVDIQVIADS
LIACVEDPVQEYKDVCKVIEKGLVENFTRAKEMAYRLNQYWFNRSVSFKIPKKI
RDPVPKDVVRQEDATTSSQSHDNSSRTVSPNHRHHSSSYNSCYQEREPSHIRFF
NNGNDVHQYRFGGYHGNNYNDNYFSRRPNKDSYRDRRRFNGRRSRSRSRVSP
QNYKRRKLDEHDNNHRQRSPIRDRHTSPGGEKTPSSNNSGERNYKRLDIRGARIK
TIKEDLEAAAAAAAAAAVPSVQAYPHEHTAVHQSVYQMPGYESYGVYDPVNG
VYMYPHPGAGYYPAYPQQPIMLTMDTLPPNDRLGELYEKASIEQLAQRDIVR
QELELIRIQIERKTAQKEAIKAACRRANEEEEAKRQEALAKTKYVWAIKSEAGET
YYYNKITKETQWTAPTPVQGLLEPACGASPDTTVVIADIEEEEEQQAQEVLEKPRV
VKEEVIEPGSQSETQKESPEKVRVVVPKVEVERSPSPKSSRDREKREKREKDR
ERDRDRREGSKHRDSYHGHRNGSSSVSERRMREFKHELERSTRSAVRSRLQHQR
DASSDKTTWLKLIYREIFKRESAQSGFDYRFSENTDKKVKKNYVKSIDRKLESN
DLWKEYSRP

Figure 25

lin(n4256) genomic sequence (1 kb of upstream and downstream genomic sequence is included in this file).

<u>Exon number</u>	<u>Exon boundaries (inclusive)</u>
1	1001 – 1096
2	1166– 1453
3	1501 – 2199
4	2298 – 2730
5	3234 – 3847
6	4148 – 5778
7	6111 – 6333

GCTTGCATCGAAACTCTTCTCATTATTTACGTGATGATCACATCTTTCGTTGGG
 CTGTACTCCCTTCCGGTTCTTCGTTCTCTTCGACCTGTTTCGAAAAGATACTCCA
 ATGCCAACGATAATTATTAATTCTTCAATAGTTCTTGTTGTTGCATCCGCTCTC
 CCAGTAGCTGTTAACACAGTTGGAATGACAACCTTTTGATCTTCTCGGCTCCCA
 CTCATCGCTCCAATGGCTTGGATCATTTTCGAGTCGTTGTTGCCTATAATACTCT
 ATTCGTCTGTGTCTGTCTCGCATTTCTCTTCAATCAATTGACTGCTTCAATGAG
 AAGGCAAATCTGGAAGTGGTAAGCTGTGCAATTTAAAGTTTAAATTCTTATTA
 ATTTTTTTGCAGGATATGTCAACTACGATGTGGAATCAGACGGGAGAGTGAT
 GCGGATGAAACCATTGAGATCCTTAGAGGCGATAAGAAAAGCAATTGAATTT
 CTTTCCTTTTTCAACACTTCTTACCCATGTTTCATCATTTTAATCTTTTCATTACA
 AAAACAAGGTCCTATTTTTTTTCTCGGGTACTACTCGCCTTTTCTAATAATTCA
 GAATCATCAATTTTTTGCCAACCTCTAGCTTTACATGTCTGTTTTTCATCATTTT
 CTCTCAAGCATTCTCCTAATATATTATGTTCCCTAGTATTTCCCCTCAGTCAGC
 AATTTTCTCGTCGTCGAAACCGTTTAGCTTTACTTTCAATCAAAACGTGGAAC
 ATTTTTCAAACCTATTTGAAGCCAAAAAAACCAGGGCTTTTGTATATGTACCA
 TATTTTCCCTCTGATTTTCTTTATCGCCTTCTCTTTTCATGTAGAATAACTGAA
 ATACAAACCATTTTAATTTTTTCTTTTAATTATCAATACTGTCCGTATAGGTAA
 AAATTATTTCTTCAGGTTTGAAAAAATCCGAAATATGTATCTGCAACTCTTCA
 GGGCATTGCCTCAATTAATTTTTATCTAATATTCAGATGGACCAACAAGAACC
 ATCGAATAACGTAGATACGAGCAGTATTCTTTCGGATGATGGGATGGAAACA
 CAGGAACAAAGTTCATTTCGTCACCTGCTGTGAGTGAAATTATTTAAAATTTTCGC
 TTCGGAGATTCAATTGTCATATAATTCAATTTATCGATTTTCAGACAATTGACC
 TAACAGTGGACGACTACGATGAAACAGAAATACAGGAGATTCTGGATAATG
 GAAAAGCAGAAGAAGGAACAGATGAAGATTCTGATTTAGTTGAAGGGATTCT
 TAACGCTAATTCAGATGTCCAAGCGCTCCTTGATGCGCCATCTGAGCAAGTA
 GCTCAAGCTCTTAATTCGTTCTTCGGAAATGAGAGTGAACAAGAAGCTGTTG
 CAGCACAAAGACGGGTTGATGCGGAGAAGACTGCCAAAGATGAAGCTGAAC
 TCAAGCAACAGGAAGAGGCGGTTAGATTGCAATAAAGGAAGCAATAATAAA
 ATTATTTTATTTTCAGGAAGATCTTATTATAGAAGATTTCGATAGTCAAACTG
 ATGAAGAAAAACAAGCAGTTCGAAGACTGAAAATCAACGAATTTTTATCGTG
 GTTCAACAAGGCTCCTTCCAGAACAATTTAAAAATTTTCAATTACAAATCCGA
 ACTATCTGACAGAATCTATCAGCGATTACCGGTTGTAAATGTGCGATAAATGC
 AAGGAAATTGTCAAATCGTTCAAGGAAAGTGAATCACTTGAGGGACTTTCAC
 AGAAATACGAATTAATTGATGAAGACGTGCTAGTCGCTGCTATTTGTATTGGC

FIGURE 25

GTTCTCGATACCAACAACGAAGAAGATGTCGACTTTAATGTTCTATGTGATGA
TCGTATCGACGATTGGAGTATAGAAAAATGTGTCACCTTTCTTGATTATCCAA
ATACTGGATTGAATTTCGAAAAATGGACCGTTGAGATTCATGCAGTTTACTGTC
ACATCACCTGCATCAGCAATTCTCATGCTCACTCTGATTTCGATTACGCGAAGA
AGGGCATCCGTGTCGATTAGATTTTGATTCAAATCCGACTGATGATTTACTCT
TGAATTTTCGATCAAGTGGAATTTTCTAATAATATCATTGATACGGCAGTCAAA
TACTGGGATGATCAGAAGGAAAACGGTGCGCAGGATAAAATTGGCAGGCGA
GTATTAATCAAACCTCACAACCTGTTTTGAAAGTATTTTCATAATTATCACTTAA
ATACCTTTTAGAGAGCTCAACGACTTCTTCCACGAAATCGAGTCAACATCAGC
AGAATTCAAACAACATTTTGAGAACGCCGTTGGCAGCCGTAATGAAATAATT
CAACTTGTCAACGAGAAAAATCCCGATTTTGATGGCACTGAGGCTGCTGTGA
ATGAGAGTTTTACATCCGATCAACGAACCGAAATTATCAACTCTCGTGCAAT
AATGGAGACATTAAGCCGAGATGAAGCTCGCCATCGCCGAAGCTCAGAA
AGTTTACGACACCAAGACTGACTTCGAAAAATTTCTCGTTTTGACAGTTGGAG
ATTTCTGTCTGGCTCGCGCCAATCCTTCTGACGATGCAGAATTAACATACGCC
ATAGTTCAGGATCGTGTGGATGCAATGACCTATAAGGTTAAATTTATCGACA
CAAGTCAGATCAGAGAGTGTAACATCAGAGATTTAGCCATGACTACGCAGGG
AATGTATGACCCGAGTTTGAATACATTTGGTGATGTTGGTGAGTTTTAAGTTA
AAATTGATATTTAATATTACATCTGTTATGTAGAATAAGGGTTTCGGTTTTTC
GATTTTATTAGAAAAATCGAAAAATTTAGTTTTTGTGTTAAATTTAAAAAAATC
AAAATTTGATTCATCAAGTCCGTTTTTCTCTTCTCAAAATTGACAAAATTT
TGATAATCTAGAATTTTTCGTCCCGTATATTTTCAACGAAAAACCATTAAAA
TTTTCCATGATTGGATTTTCGGTTGATCTAGAAAAAAATGGTGCTAAACACTA
AATTTGAAAAAGTTTGAAACAAATTCAAATCCAAATATTTTCATGAAAACTT
GTAAAATATATTATGTACACAAAAAAACGTTTCAAGTGTAGCAGTTGTTTTTT
GTGGTCCCAAAAAAGCAGATGTTTGTGAGAATCCATTAAACAACAAAAAAAT
CCAAAAACTCAACCTGGCCTAGATATCAGTTTCATGATCGAAGTATCTAAAA
TCATTGTTTTCAGGTCTTCGAGTTGCCTGTCGCCAAGTTATTTCTCGAGCCAA
TTTGAAAAAAAACAATTTGGCTTACCGGTACAGCTGCCGGACGTCGCAGAG
CTCATAGATCCGATTTTCTAATTTTCTTCGACAACCGGAACCGATGCATACGTG
TCAGCTCCGACAATGCCTGGTGAACCAGGTTATGAAGTTGCTTCTGAAAAGA
AAAGTGTATTTTCTCTCAAAGAAATGATTGCGAAGATGAATGCTGCTCAGATT
GCTATTATGGTTGGACAGCCAGTAGGAAAGGAAGGAAATCTGGATTATTTTT
TGACATTTTCATTGGATTTCGACAATCTCACAGATCAGCGTATATTCGGGATTTT
ATGAAAGAATTTCCGGAATGGCCACTTCTCAAGATGCCAGTTGGAATGCGAA
TCTGTTTGTACAATTCTCTTGTGATCGACGTAAGAAAAATGGTGACAGTGATT
GGAAGTATCGAGCTTTTGCTATTGTGAGACACGAAGCACCGAATCCATTGG
CTCCTGGGAATAGATGTACAGACTTCCGTGCAATGATAGAAATCATCAGCA
TATTGACGAGAAAAATCTATAGAGGATCTCATAGATTGGAAGGCGCAGCGGTA
AGATTTTATTTGAAAAATTGATACAAAACGAGGATTTTCTAAAATTATTTTAT
TTTTATTGATTTGATTTCTTATAATTGATAATCAAGGTTTTTTGGATGTTTTG
TTAGAGAAATCGAAAAGGGAACTTCCAAAAAAAAGCTGTGAAATCAATTTT
TGCTTTTAATAATATCCAAGTTTCATCTTCAAAGTTTTTTCTATAAAATGGACA
CAAACCTTTCAACGTTTTTCAAAAAAAAGGTTCCGAAAATATGAAAAAGGAG
AAAGAAATCATGAAAAATTTGTATTATTTTCAGCACAGAAGCACATGATCTC
GACAAATAACAATCTGTCGCAACGCAGAAAAGACCAGCTTCAATCACAGTTC
GAGCCAACCGACATGATTTCGTTTCGATGCCAGAGAGGAATCACCAACAAGTCG
TAAAAAAGAAAACGACGGGCACCAATCAGAATGTCGCTTCGACAAATGATGC

FIGURE 25

AAAATCGAAGAGAGAAATTGAAATAAGAAAGAAAAATCAATTCTTATTTAAC
AAGATTATTGTTCCAATACCCGTCCTAACACCATTGGAAAATCTCAAGGCTCA
TGCTCAATGTGGTCCAGATTGTCTACAGAAAATGGATGCGGATCCGTATGAA
GCAAGATTCCATCGAAATTCACCAATACATACTCCTCTTTTGTGTGGTTGGAG
ACGAATTATGTACACAATGAGTACTGGAAAGAAGCGGGGAGCAGTGAAGAA
AAACATTATTTACTTTTCTCCATGCGGAGCCGCTCTTCACCAGATCAGCGACG
TCTCTGAATATATTCATGTCACCAGAAGTTTATTGACGATTGATTGTTTTTCAT
TTGATGCACGAATCGATACTGCCACTTATATTACTGTTGACGATAAATATTTG
AAGGTTGCTGATTTTTTCGCTTGGAACCGAAGGAATCCCAATTCCACTAGTGAA
CAGCGTGGATAACGATGAGCCTCCATCATTGGAATATTCGAAACGACGATTC
CAATACAATGATCAAGTGGATATATCGAGTGTTAGCCGAGATTTCTGTTCTGG
ATGCTCTTGATGGTGATTGCAGTGACGCATCGAAGTGTGAATGCCAACAA
TTGTCCATTGAAGCAATGAAACGACTCCCCATAATTTACAATTCGACGGAC
ACGACGAATTGTATGAGAGTTTCAAGAAAACAAAATAAATTTTTTAAACTATT
TTTTTTCAGAGTTTCTCACTATCAAAATCGTCTTCTCAGCAGTAAGGTTATCA
GTGGACTCTATGAATGCAACGATCAGTGTTTCATGCCATCGAAAGTCTTGTTAC
AACAGAGTTGTTTCAAGACAATATCAAGTATCCTATGCATGTGAGTTTATTTAA
CGATGATACATACCAATTATTGTTTTTCTTCAGATCTTCAAACTGCTCAATC
CGGATGGGGAGTCCGAGCTTTGACGGATATTCCTCAAAGTACGTTCAATTTGCA
CGTATGTAGGTGCTATACTGACGGATGATTTGGCTGATGAACTAAGAAATGC
GGATCAATACTTCGCTGATTTGGACTTGAAGGATACCGTGGAGCTGGAAAAG
GGTCGCGAAGATCATGAACTGATTTTGGTTACGGAGGAGACGAGTCAGATT
ATGATGACGAAGAAGGAAGTGATGGTGACTCCGGTGATGATGTAATGAACA
AAATGGTGAAACGTCAAGACTCTTCGGAGAGTGGTGAAGAAACAAAACGGC
TGACAAGACAGAAAAGAAAGCAATCTAAAAAATCCGGTAAAGGAGGAAGTG
TGGAGAAAGATGACACCACTCCAAGAGATTCAATGGAAAAGGATAATATTG
AAAGTAAAGACGAACCCGTTTTCAATTGGGATAAGTATTTTGAGCCGTTTTCCA
TTGTATGTTATAGATGCAAAACAGAGAGGAAATCTTGGAAGGTAAGATCACA
ATTTTATTCATTAAAAAAATTTTTAGAGATTTTGCTTTAAATGATAAAAAAT
GGACAAACCAACCGTTTTGCCTCTTCTTTTGGTTTATCAACCTTTCTCTATGGAA
AAAATTCTGAAAAATTAACAAACAGTATTTACGTTGAAAAGTGAAGAAAAA
AGCAAAAAAAGGAAACAAATTTCAAAACGGTTCTACTCCATCTTAAAAAAAC
TAAATTCGTAAAAAGTCATTTGGTATGTTTTGGAGACTATAATACAATTGAG
AAAATTTGAAAAACCGGCACTCCAAGATACAATCATAAATTTTCGATAACT
TTCAGATTCTTGAATCACTCTTGCGATCCGAATGTGCACGTTCAACACGTCAT
GTACGATACGCATGATCTTCGTCTTCCATGGGTGCGGTTTTTCACACGAAAAT
ACGTGAAAGCCGGCGATGAGCTAACCTGGGACTATCAATATACTCAAGATCA
GACGGCTACCAACAACCTACATGCCACTGCGGAGCTGAAAACCTGCACCGGC
CGTTTGCTGAAAAGTTAAAGAATTGTTGTTATTTCTTCCCAGTTATGTTTTCC
TTTTTTTTTAAGTATTTATTTATTTAATTTTTATTTTGTTTATTGTTCAATC
GTTTAAAATCTCCCTTTGAAAACAGCATCTCATATGTATGATCTAAACACGTA
TTTACCTCGTAAGGGTTTGCCAAATAGTTTCTTTGGTTTTTCATTTTGATTTTCT
CTGCGAATAAAATGTTTTAAAAAAGACATTATTTTTTTAATAGTCAGTACAG
TTTTGATGTCTCCAATCTATTTTCAAGTTTACAATTTTAAAAATATAGAATATATAT
ATTTAGGTTTCATAAGTTATGCATCGATTACGGGTTCTACGTCACTTGAAGTT
CTGCATTTCCACGTCACATAGGACTACTGTAGTTTTTAAAAAATACTCGTTCAT
TTTGTAATAATATTCCTTCTACTAGTTTTGCTTCTGGTAATAATCGAATTTCAA
AACTTTAGCTAAAAATATTTCTTTTTGAAGAGGCTGCAGCAAAATATGAAAAG

FIGURE 25

AAAAGTCCAACTGAACATGTATTACTTCGACCCGATACATATATTGGAGGTG
TCGCCATGCGAGAAGATCAAATTATTTGGCTCAGAGACTCAGAAAATAGAAA
AATGATTGCAAAAGAAGTCACTTATCCACCTGGATTATTGAAGATTTTCGATG
AGATTCTAGTGAATGCGGCTGATAATAAAGCAAGAGATTCCAGTATGAATCG
GTTGGAAGTATGGTTAGATAGGTAAATATATTGCAGGAATTTATGTTCTGCGA
CAAAGCTACGATACGCTGTCTCGCCACGACAATTGTTTTGGTAAATGCATGA
AAATCGACGTGCACCTTTAAATAATACTGTAGTTTTAAATTCTCGTTTCTTCA
ATTTTTCATAAATGGTTTTCCGATGAATATATGATTTTAAAAAATCTAAAT
TCACATTAATTTATAAGAAACAAAATTCCTCAAAAACGAAAGTTTGGCGATA
CAGTACTATC

Figure 26

LIN(n4256) amino acid sequence

MDQQEPSNNVDTSSILSDDGMETQEQSSFVTATIDLTVDDYDETEIQEILDNGKA
EEGTDSDSLVEGILNANSQVQALLDAPSEQVAQALNSFFGNESEQEAVAAQRR
VDAEKTAKDEAELKQQEEAEDLIIEDSIVKTDEEKQAVRRLKINEFLSWFTRLLPE
QFKNFEFTNPNYLTESISDSPVNVNVDKCKEIVKSFKESSESLEGLSQKYELIDEDVL
VAAICIGVLDTNNEEDVDFNVLCDDRIDDWSIEKCVTFLDYPNTGLNSKNGPLRF
MQFTVTSPASAILMLTLRLREEGHPCRLDFDSNPTDDLLNFDQVEFSNNIIDTA
VKYWDDQKENGAAQDKIGRRVLIKLTTLKNAVGSRNEIIQLVNEKIPDFDGTGA
AVNESFTSDQRTEIINSRAIMETLKAEMKLAIAEAQKVYDTKTDFEKFVLTVDG
FCLARANPSDDAELTYAIVQDRVDAMTYKVKFIDTSQIRECNIRDLAMTTQGM
DPSLNTFGDVGLRVACRQVISSSQFGKKTITWLTGTAAGRRAHRSDFLIFFDNGT
DAYVSAPTMPGEPGYEVASEKKSVSFLKEMIAKMNAAQIAIMVGQPVGKEGNL
DYFLTFFHWIRQSHRSAYIRDFMKEFPEWPLLKMPVGMRIKLYNSLVDRRKKMVT
VIGTDRAFAIVRHEAPNPLAPGNRCTDFPCNDRNHQHIDEKIYRGSHRLEGAAHK
KHMISTNNNLSQRRKDQLQSQFEPTDMIRSMERNHQQVVKKKTTGTNQNVAS
TNDAKSKREIEIRKKNQFLFNKIIVPIPVLTPLENLKAHAQCGPDCLQKMDADPYE
ARFHRNSPIHTPLLCGWRRIMYTMSTGKKRGAVKKNIIYFSPCGAALHQISDVSE
YIHVTRSLLTIDCFSFDARIDTATYITVDDKYLKVAADFSLGTEGIPIPLVNSVDNDE
PPSLEYSKRRFQYNDQVDISSVSRDFCSGCSCDGDASKCECQQLSIEAMKRL
PHNLQFDGHDELYESSEKQNKFLKLFFFRVPHYQNRLLSSKVISGLYECNDQCSC
HRKSCYNRVVQNNIKYPMHVSLEFNDDTYQLLFFLQIFKTAQSGWGVRAITDIPQ
STFICTYVGAILTDDLADLRNADQYFADLDLKDTELEKGREDETDFGYGGD
ESDYDDEEGSDGSDGDDVMNKMVKRQDSSESGETKRLTRQKRKQSKKSGKG
GSVEKDDTTPRDSMEKDNIESKDEPVFNWDKYFEPFPLYVIDAKQRGNLGRFLN
HSCDPNVHVQHVMYDTHDLRLPWVAFFTRKYVKAGDELTWDYQYTQDQTATT
QLTCHCGAENCTGRLLKS

Figure 27

lin-65 genomic sequence (1 kb of upstream and downstream genomic sequence is included in this file)

<u>Exon number</u>	<u>Exon boundaries (inclusive)</u>
1	1001 – 1133
2	4522 – 5208
3	6128 – 6361
4	7962 – 8350
5	8706 – 8928
6	9260 – 9516
7	10328 – 10567
8	11677 – 11700

```

AAAAATTTAAAAAAATTTTTAAAAATTCGTGTAAAAATTACCCCGGTTGTTTA
GGAAATAATAAAGAGATTAGAGACTTTTTTCAGATTTTATTTTCTTGAGTTT
TGCCGGTTTTTCAGCCGATTTCTATCTTTTTTTTCTCATTTTTTGTGATTTTTTT
CGCTAGTTTTCCCTCAATTTCTCGATTTTTTCACGATTTTTTGAAAATTTTCG
GAAAATTGAATTGTTTGCAAAAAAAAAAAATTCAAAAACCGCATTTTTCTCAG
AATTTTTCTGGGATTTTGTACAAATTTTGAATTATTTCTCAAAAAAAAAAGCAG
GTTTTTACCGATTTTTTTGGTTTTTCCCCAAAATTTTCCGATTTTTTCCGAGTT
TTGCCGGTTTTTCAGCCGAATTCTACTCTCGATTTTTTTTACGATTTTTTGGAAAT
TTTCGGAAAATTATTTGAAAAAAAAATCAAAAAACCGCATTTTTTTTTCTGAAT
TTTCTGGGATTTTGTACGAAATTTTGAAATTTTTCTCGAAAAAAGCAAGTTAT
TCCCCAAAATTTTCTGATTTTCCCCCAAAAATTTAGATTTTTTCCCGAGTTTTCC
CCAGTTCTCAGCTGATTTCTATATTTTTTTTCTCAATTTTTGTGATTTTTTGTGTC
TAGTTTTCCCTTCAATTCCTCGAGTTTTTCACGATTTTTTGGAGATTTTCGAAA
AATTGTTTGAAAAAAATCAAGAAACCACATTTTTTCTCTGGATTTTCTCGAAAT
TTGCACAAAATTTTTGAATTTTTTCGTAAAAAAAAAACTGTTTTCCCCAAAAT
TTCAGATTTGTTTTTGATTTTTTTTCGAGATTTTCCCTGATTTCAAAGTTTTTTC
CTGAATTTTTTCGAATATTTCTGAAAAATCGGCTATTTCTAACTTTTTAAATAA
TTTTTTTTGAATTTCTGACTTTTTAAATCCTTTTTTTTTTGCCATTTTTTCCCATC
TAAATTTCTAAATTATTCAAAATTTTACAGAATGTCAGAAGTAATCGACGAA
AGTATCTTAAATACAGAAGCTTCAGATGATCCAATACCTCCATTAAATGATG
ATCAGATTGCTGAGCTTTTGGGTGAAGATGGAGAAATTATGGAGATAACTGA
GCAGAAAGGTGAGATTTTTTGAGTAAAACCTTGAATTTTGCCTAAAAATTTG
CAATTTTCGCTAAAAATTACCTTAAAACTCGAAAATTGGAATTTCTAGCTGAG
AAAATGGCCAAAAATGTCGAAAAATGCCTCCGAAACCTGTGAAAAAAAAAAAA
CCACCAAAAAGGTTTCTAGGCCACCAAAAAGATTTCTAGGCCACCAAAAATG
TTTCTAGGCCACCAAAAATGTTTCTAGGCCACCAAAAATGTTTCTAGGCCACC
AAAAATGTTTCTAGGCCACCAAAAATGTTTCTAGGCCACCAAAACAGGTTTCA
ATGCCACCAAAAATGTTTCTAGGCCACCAAAAATGTTTCTAGGCCCCCAAAA
AATTTTTCTAGGCCACCAAAAAGGTTTCTAGGCCACCAAAAATGTTTCTAGGC
CACCAAAAAGGTTTCTAGGCCACCAAAACAGGTTTCAATGCCACCAAAAAGGT
TTCTAGGCCACCAACCAGGTTTCAATGCCACCAAAAATGTTTCTAGGCCACCA
AAAAGGTTTCTAGGCCACCAAAAATGTTTCTAGGCCACCAAAAATGTTTCTA

```

FIGURE 27

GGCCACCAAAAAGGTTTCTAGGCCACCAAAACAGGTTTCAATGCCACCAAAAA
TGTTTCTAGGCCACCAAAACAGGTTTCAATGCCACCAAAAAGGTTTCTAGGCC
ACCAAAAAGGTTTCTAGGCCACCAAAAATGTTTCTAGGCCACCAAAAAGGTT
TCTAGGCCACCAAAACAGGTTTCAATGCCACCAAAAATGTTTCTAGGCCACCA
AACAGGTTTCAATGCCACCAAAAATGTTTCTAGGCCACCAAAACAGGTTTCAA
TGCCACCAAAAATGTTTCTAGGCCACCAAAAAGGTTTCTAGGCCACCAAAA
TGTTTCTAGGCCACCAAAAATGTTTCTAGGCCACCAAAAAGGTTTCTAGGCCA
CCAAACAGGTTTCAATGCCACCAAAAATGTTTCTAGGCCACCAAAACAGGTTT
CAATGCCACCAAAAATGTTTCTAGGCCACCAAAAATGTTTCTAGGCCCCCAA
AAAATTTTCTAGGCCACCAAAAAGGTTTCTAGGCCACCAAAAATGTTTCTAG
GCCACCAAAAAGGTTTCTAGGCCACCAAAACAGGTTTCAATGCCACCAAAAAG
GTTTCTAGGCCACCAACCAGGTTTCAATGCCACCAAAAATGTTTCTAGGCCAC
CAAAAAGGTTTCTAGGCCACCAAAAATGTTTCTAGGCCACCAAAAATGTTTCT
TAGGCCACCAAAAAGGTTTCTAGGCCACCAAAAAGGTTTCAAGGCCACCAAAA
AAGGTTTCAATGCCACCAAAAATGTTTCTAGGCCACCAAAACAGGTTTCAATG
CCACCAAAAAGGTTTCTAGGCCACCAAAAATGTTTCTAGACCACCAAAAAGG
TTTCTAGGCCACCAAAACAGGTTTCAATGCCACCAAAAAGGTTTCTAGGCCAC
CAAACAGGTTTCAATGCCACCAAAAATGTTTCTAGGCCACCAAAAAGGTTTCT
TAGGCCACCAAAAATGTTTCTAGGCCACCAAAAATGTTTCTAGGCCACCAAAA
AAGGTTTCTAGGCCACCAAAACAGGTTTCAATGCCACCAAAAATGTTTCTAGG
CCACCAAAACAGGTTTCAATGCCCCCAAAAATTTTCTAGGCCACCAAAAAG
GTTTCTAGGCCATCAAAAATGTTTCTAGACCACCAAAAAGGTTTCTAGGCCAC
CAAAAATGTTTCTAGACCACCAAAAAGGTTTCTAGGCCACCAAAAATGTTTCT
TAGGCCACCAAAAAGGTTTCTAGGCCACCAAAAATGTTTCTAGGCCACCAAAA
AAGGTTTCTAGGCCACCAAAACAGGTTTCAATGCCACCAAAAAGGTTTCTAGG
CCACCAACCAGGTTTCAATGCCACCAAAAATGTTTCTAGGCCACCAAAAAGG
TTTCTAGGCCACCAAAAATGTTTCTAGGCCACCAAAAATGTTTCTAGGCCACC
AAAAAGGTTTCTAGGCCACCAAAAAGGTTTCAAGGCCACCAAAAAGGTTTCA
ATGCCACCAAAAATGTTTCTAGGCCACCAAAACAGGTTTCAATGCCACCAAAA
AGGTTTCTAGGCCACCAAAACAGGTTTCAATGCCACCAAAAAGGTTTCTAGAC
CACCAAAAAGGTTTCTAGGCCACCAAAACAGGTTTCAATGCCACCAAAAAGGT
TTCTAGGCCACCAAAACAGGTTTCAATGCCACCAAAAATGTTTCTAGGCCACC
AAAAAGGTTTCTAGGCCACCAAAAATGTTTCTAGGCCACCAAAAATGTTTCT
AGGCCACCAAAAAGGTTTCTAGGCCACCAAAACAGGTTTCAATGCCACCAAAA
ATGTTTCTAGGCCACCAAAACAGGTTTCAATGCCCCCAAAAATTTTCTAGGC
CACCAAAAAGGTTTCTAGGCCACCAAAAATGTTTCTAGACCACCAAAAAGGT
TTCTAGGCCACCAAAAATGTTTCTAGACCACCAAAAAGGTTTCTAGGCCACC
AAAAATGTTTCTAGGCCACCAAAAAGGTTTCTAGGCCACCAAAACAGGTTTCA
ATGCCACCAAAAATGTTTCTAGGCCACCAAAAATGTTTCTAGGCCCCCAAAA
AATTTTCTAGGCCACCAAAAAGGTTTCAATGCCACCAAAAATGTTTCTAGGC
CACCAAAAAGGTTTCTAGGCCACCAAAAATGTTTCTAGGCCACCAAAAATGT
TTCTAGGCCACCAAAAAGGTTTCTAGGCCACCAAAACAGGTTTCAATGCCACC
AAAAATGTTTCTAGGCCACCAAAACAGGTTTCAATGCCACCAAAAAGGTTTCT
AGGCCACCAAAAATGTTTCTAGACCACCAAAAAGGTTTCTAGGCCACCAAAAC
AGGTTTCAATGCCACCAAAAAGGTTTCTAGGCCACCAAAACAGGTTTCAATGC
CACCAAAAATGTTTCTAGGCCACCAAAAAGGTTTCTAGGCCACCAAAAATGT
TTCTAGGCCACCAAAAATGTTTCTAGGCCACCAAAAAGGTTTCTAGGCCACC
AAACAGGTTTCAATGCCACCAAAAATGTTTCTAGGCCACCAAAACAGGTTTCA

ATGCCACCAAAAATGTTTCTAGGCCACCAAAAATGTTTCTAGGCCCCCAAAA
AATTTTTCTAGGCCACCAAAAAGGTTTCTAGGCCACCAAAAATGTTTCTAGAC
CACCAAAAAGGTTTCTAGGCCACCAAAAATGTTTCTAGACCACCAAAAAGGT
TTCTAGGCCACCAAAAATGTTTCTAGGCCACCAAAAAGGTTTCTAGGCCACC
AAAAATGCTTCTAGGCCACCAAAAATGTTTCTACGCCACCAAAAAGCCGCCTC
AAGCCCGAAAAATTTGAATTTCCCGCTCAAAAAATCTAAAATTTTCCGATTTT
CAGACGAATCAGATGATGTGGTGATGCTGGACGACGATGATGACGACACTCC
GGAACCGATTCTCGTGATTGATATGGATGAGGATGAGGATGTTACTACAGAT
GGTCCTGAATCTCAGGAAGAGCTGGCTGCAGATGCTCCGGCTCCAGGAGCTC
CAGAAGCTTCAGCTCCAGCTCAAGAAGCCTCAGAAGCTTCAGCTCCGGATCA
AGAAGCTCCAGAAGTTCAGGATGTTCCGGATTCTTCGGGAGCTCCAGATGCT
TCAGCTCAGGCTTCAGAGGCTTCTGATGCTTCAGCTCCAGAAGTTCAGGATC
TACAGAAGCTCAGGATGCTCAGGATGTTCCGGATTCTTTGGGAGCTTCAGAT
GCTTCAGCTCAAGAAATTCAGAAGCTCCAGAAGCCCCAGAAGCTCCAGAAA
TCGCCGCTGAAATCGACGAAGAAGTGCTGCTCGCCGAGCAAAATGGAGTTTT
GGACGAAGGATTTGATGAGACTGACGATATTATCATAGAAGAAGAAGCTGTA
GAAGAAGCTGAAGCCGTGGAGCCACCAATTAACACTGAAAATCAGGAAAAC
GCGCTGGAAATGCTCGAAGAGCGCCTCAAGAAGAATGAAGAAAAGGAAATT
GTGGAGAAAAGTGATGTGAAGCCAGAGGATGAAGATATTATACATATGGAG
ACGGATTCAGTTGAAAGTATGGGCTTTTTTAGCTGGAAAACAGGAAAAAAGA
GCAAAAAATTGATACATTTCCAGCTTAACCAATCTTTTTTTGAGTTGTAAAGC
CTGAAAATTGAGATTTTTGTACCAACTTTTATGATAAAGCTGAAAAAAAATT
AATTTTTTGACGAATTTTTAGCGGAAACCCTGAAAACATGTTTTGTCTGAAAA
ATACAGAAAATCGTCACTTTTTACAATAAATTCGAGATTTTTAGCTCAAAAAT
ACAACATTATAGTGCAAAAATCTCAGAAAAAGCCAAAAATTTCAATCAACA
TCTCAAAAAAAGCAGAAAATTTTACTCAAAATATCTCAGAAAAAGCTAAAATT
TTCCCAAAAAATCCCAGAAAAAGCAGAATTTTCAATCAAAATTTCCAGAAAA
AGCTGATAATTTACTAAACAATCTCAGAAAATGCTGAAATTTTACTCAAAAG
TCTTCATAAAAAAGCTGAAATTTTACTTTAAAAGTTTAGGAAATGCTGCAATTT
CACTTAAAAATCCCAAAAAAGCTAAAATTTTCCCAAAAAATCCCAGAAAAAG
CAGAAATTTTACTCGAATATCTCAAAAAAAGCTGAAATTTCACTCAA
AAATCCCAGAAAAAGCTAAAAATTTACTAAAAAATCTCAAAAAAAGCG
CTAAAATTTCACTCAAAAATCTCAGAAAAAGCTAAAATTTTACTCGAATATCT
CAAAAAAAGCTGAAATTTTCTTAAAAAATTTATGAAAAACCGAAATTTT
ACTTAAAAGTCTCATAAAAAGCCGAATTTTCCCAAAAAAATCCCAGAAAAAG
CTAAAATTTACTTTAAAATCTCATCTGTAATTTTAGTTTAAAATCTCAGAAA
AACCCGAAATTTCTCTCAAAAATTTGCTGATTTTCAAATTTTCAGCGTCAAGC
CGCAAACGTAAGTGGCGGAGCCACAAGTCCGCGGAGCCCGGCTCAAAAACGA
CCAAAACGACGTGTTCAAACGTTATTAAAGATGCGTCAGAATGCAATTGAAC
TATTGACACGACTTTATGGCTCATGGGATGCACAATTGAGCCTCTCAAATCTT
GAGACAATTCGATTGTTGGGTGTCAATAATAATAGGAAGCTTATCGAAATTTT
TGAGGAGAATGAGCAAGGTTAAAGCGTTTTTAAATGCTATGAAAACCTGACAA
ATTTTCGATAAAAAAAGCGATTTTGGGAAGAAAATCGCCTGAAAATTCATGT
TTTTCTGCAAATTTTGACCAAATTTCCCAAGAAAAATACGATTTTTTAGTCCGA
AAATCCTCCAAAAAGATTTCTAGGCCACCAAAAAGGTTTCTAGGCCACCAAG
AAAGTTTCTAGGCCACCAAGTATTTATAGGCCACCTAAGATGTTTCTAGGCC
ACCTGAGATGTTTCTAGGTACCAAAAATGTTTCTCGGTACCAAAAATGTTT
CAAGGCCACCGAAAAGGTTTCTAGGCCACCTAAGTATTTCTAGGCCACCTAA

FIGURE 27

GATGTTTCTAGGCCACCTGAGATGTTTCTAGGTCACCAAAAATGTTTCTAGGT
TACCAAAAATGTTTCAAGGCCATCGAAAAGGTTTCTAGGCCACCAAAAGTATT
TCTAGGCCACCTAAGATGTTTCTAGGCCACCTGAGATGTTTCTAGGTCACCAA
AAATGTTTCAAGGCCACCGAAAAGGTTTCTAGGCCACCAAAAAGGTTTCTAG
GCCACCAAAAATATTTCTAGGCCACCTAAGATGTTTCTAGGCCACCTGAGAT
GTTTCTAGGCCACCTGAGATGTTTCTAGGCCACCTGAGATGTTTCTAGGTCAC
CAAAAATGTTTCTCGGTACCAAAAATGTTTCAAGGCCACCGAAAAGGTTTCT
TAGGCCACCTAAGTATTTCTAGGCCACCTAAGATGTTTCTAGGCCACCTGAGA
TGTTTCTAGGTCACCAAAAATGTTTCTAGGTTACCAAAAATGTTTCAAGGCCA
TCGAAAAGGTTTCTAGGCCACCAAAAGTATTTCTAGGCCACCTAAGATGTTTCT
AGGCCACCTGAGATGTTTCTAGGTCACCAAAAATGTTTCAAGGCCACCGAAA
AGGTTTCTAGGCCACCAAAAAGGTTTCTAGGCCACCAAAAATATTTCTAGGC
CACCAAAAATGTTTCTAGGTCACCAAAAATGTTTCTAGGTCACCAAAAATGT
ATCAAGGCCACCAAAAAGGTTTCTAGGTCACCAAAAATGTTTCTAGGCCACC
AAAATGTTTCTAGGTCACCAAAAATGTTTCTAGGCCACCAAAAAGGTTTCT
AGGCCACCAAAAAGGTTTCTAGGCCACCAAAAAGGTTTCTAGGCCACCAAAA
AGGTTTCAAGGCCACCAAAAAGGTTTCTAGGCCACCAAAAATGTTTCTAGGT
CACCAAAAATGTTTCTAGGCCACCAAAAGTATTTCTAGGCCACCTAAAAGGTTT
CTAGGCCATCAAAAAGGTTTCTAGGCCATCAAAAAGGATTCTAGGCCACCAA
AAATATTTCTAGGCCACCTAAGATGTTTCTAGGCCACCAGAGTATTTCTAGGC
CACCTAAGAGGTTTCTGGGCCATCAAAAAGGTTTCAAGTCCATCAAAAAGGT
TTCTAGGCCACCAAAAAGGTTTCTAGGCCACCGAAAAGGTTTCTAGGCCACC
AAAAGGTTTCTAGACCACCTAAGACATTTCTAGGCCAACAAAAAGGTTTCT
AGGCCACCAAGAAGCCGAAAACTGTCTCAAATTCGAATTTTGCAGTGCTCA
AACAAAAAGTGTCGCACTGACAGAAGAGCTGAAAAAGGAGAAGCTGGCTC
ACGCGGGAACCCGTTCAAGCATTGAAAGAATTGACTAATGAAATAACTGGAAT
GCGTGTAACAATGAATAAACTACGTTCAATGGTCACTCAGCCTACGACTTCG
AAAATTATTGATAGTTTTGTTCAACGTCATCAGGCTTTCGAGCAGCAACAACA
ATTCCAACACCAACACCACCAACACCGACCAATAATGTTGGCTCCACGTCAT
CATCCGCCGCCGCCCGCATTTTACACCGAATCAACGGGCGGCGGCTCCGT
ATCATCCGAATATGGTTCAACCGAATCGTCTTGCTGCTATGCCACATAGAAGA
CCGATTATTGGAATGCAGGTGAAAATGGAATGCCATGAAAATTTCCGGGCCGG
AAAATTTTGGAATACTCTAAATTTTCAATATTTGTCGAAAAAATCTGACAA
AAATCGTGTCAAAATTCAGATTTCCGGGAGAAAAATCGCATTTTGTAGTAAA
AATTCGAAGAAAAGCGTCTTAAATTCTAGATTTATTAGTTAAAAATTTTTTCA
AATTTAGTCAAGAAAATTAAGAAAAATGCGAAAATTTTCGAGCAAAAAATAT
AGTTTTTTGGAGCCGAAATTGTGAAAAATGCGATTTTTTTTCGAAAAATCTGGA
CAAAAAATTTCAAACAAGAAAAACCCTTTTTTAAAAAAATTTTCACACAAT
TTCCAGCAACAAAATTCGGCTCCACCACAATTCAACGGTACCAAGCTCTCGT
CCCATCACCTCAATCATCATCTGCATTTTCTCGTCCACCACCAACTCAACTTG
CAACACAGAGAAGAGCTCCACCATTGGCAAGTACCGGCCTTCCGGCAACAGT
CAGATGGGAAGCAATTCCACCGCCAAAAAATCCGAATGTCGGGCACAATGA
GCCACCGCTTAACAATGGAGGTTTCGTGCTGTGCAACAAAAAGAGCACCGCTT
TTCCACGACGAGTTTTTGCGATGATGATTTTGGTGTGAAAATTGAAAACTCA
TTTTTTTAAAGTCTGAAATTTGAAAATTTGAGAAAAGTTTTTTAAAAAAAGTT
TTATGAGGGATTTTCTGACAATTTTTTATAAACGGAAAATTACGAAAACCTCCA
AAATTTGTGTTCTTTTCGGAAAACGAATTTGAAATTTGAACCAAAAATTTTGACA
ATTTTCTGGGGATTTTTGACTGGAAATTCGTTTTTTCATCGATTTTTCTCCTTT

FIGURE 27

AATTTTCGGTAAAACCCCTGTCTCCAATTCCAGGCCGTGCACAGCCACTAATC
GATAATACACGTGTACACGACAATACAATTATGCTGTGTGTACCACTTGTCTC
CACTGCAAATACAATATCATCGGGCGATTTCGACACGTCTACCAAAAAGTACCA
CGAATCTACGAGAATCTCACGGCAAATCCCGATTGTGAGTGTGACGATTCATTC
GAGTGCACAGGATTTCCGAGAGAATTATCAAATTGGTGGAAAGATTAAGTAT
GAATATCTCGGAGGATTTGATCAATATGTAGGTGATGATGTTTTTTTTATTGAG
AGATAAATACGAAATTCATTACAATCGATATTTTTTGACTGAAAAATGTCTG
AAAAATCAAAAATTTTAGCTAAAAATTGAGAATATTTTTGTTTAAAAAAAT
CATTGAAATTGATTTTTTTTTATTCCATAAAAAATCTCGGAAAAGTCAATTTTC
AGTCATAAATCTTCTGAAAATTATCCAAACAATGGGATTTTCTGAAATTTTAG
CTTAAAAATTGAGGATTTCCCGGTTTTTTCAGAGAAATTCCATTACAATCGAT
TTTTTTACTGAAAAATCCTCTGGAAATTAACAAAAACCAAATAAAATGCCCT
AATTTTTTTTTTAAATCCAAAAATTGTTGGATTTTTTCAGAAAAAAATATTTTT
CAATTGACTGGTGTCCAAAAATATAGAAAATTCAAATTTTCCAAGAAAAAT
AGCCAAAAAAATGTAATTTTTGTCTAACAAAAAAATTGAATAGCGCAAAAT
AAATTGTCGTTTTTTTTAATTTCCCTCCGGTTTTGAAAGGAAAAAATTCATA
AAAATCGAAATTTTTTGACTGAAAAATCCATGAAAATCGAATTTTGAGTCA
AAAATCCTCTGAAAATGCTCCAAAAATATGAGATTTTCTGAAATTCATCAAAA
ATTAAGAATTTACGGTTTAAAAAAATTCATTAAAATCGATATTTTCAAG
TGAAAAATCTCTGGAAAATCGATGTTTGAGTCAAAATTCGTCTGAAAATGC
TCCTTTAAATTGAAAAATTGAAAAAAAACCGCCACAATATTTGCAGAATA
TCCAAGTGTTTCGTCCAAGTGTCATCTCTTAAATTCAGTGAATGAACGGTTAC
CCGGATCCAGAAGATCGTATATCAATTGACTGGGGATGCTCGAAATTGTGGC
CTTGTAAGCCGAAATCTCATCAAAATTCGGTGTACGCTTCCATCAAGCACAA
CTGCTGCCGAAGAACGATCGAATTACGATTGTGGCTGTGGCGAAGGATAAAA
CTAGCGGAATTATTCACATTTTCGCAGGTGAAAAATTGGAAAATTTGCACAAA
TCCAGACAAAAAAACTGAAAAATCGAAAAAATTTTTGTAATTTTTTGCCGA
AAACGAAAATTAAAAACTGATAAAAATTGATTTTTTAACCGGAAAATCCCTGA
AAAATCAACATTTTTTGCTAAAAATTGAGAATTATACGGTTTTGGGTAAAA
AAAAACTATTTAAAAAAATATTTTTCTTTAAAAATCTCAACAAAAAAA
ACCAATTTTCATTGAGAAATCCCCCGGAGAATTGTCAAAATTTTGGGAATAC
TCTGAAATTTTCGATAAACACCTCATTTTTGATFAAAATTGATTTTTTAAGTGA
AAAATCCCTTAAAAAACGAATATTTTAGTTTTTCACAAAAAAATGTGCAATT
TATCTGAAATTTTCAGCAAAAAAATGAAAAAATTCAGAAATTAATAA
ACTGATAAAAATCGATTTTTTACTTGAAAAATTCGTGAAAAATCAAACACATT
TTTGCTAACCATTTGAGAATATTACGATTTTGTGAAAAAACCATTAAAA
TTGATTTTTTATTCCTAAAAAATGCCAGAAAAATCAATTTTCAGTCAAAAATC
ACCGGAAAATTATCAAAATTTTGAGGTTTTCTGTGAAATTTCAAGCTGAAATT
TCCATTTTTGAATAAAAAAATGTGGCTGGATTTAAAAAACCATTAAAA
TTGATTTTTTAAGTAAAAATCCGTATTTCTCTGAAATTTTCAGGCAAAAAATG
TCATTTCCGAAATTAATAAATTGCGACAAAATCAAATAAAATTGATCAAATTT
GCAAAAAAATAAATACTTTTCGAAAAAATCCTTAAAAATTACATTTTGAAC
AAAAACTCGAATTTTCAGTCAAAAATTCGTCTGAAATGCTCCAAAAATATGG
GATTTTTTGAAATTTTAGCTAAAAATTGAGAATTGCACGGTATTTAGAGAGGG
AAAAATTCATAAAAAATCGATATTTTCTTTTAAATCTCGAAAAAATCAT
CAATTTTCATTCAAAAATCCCCCGGAAAATTGTCAAAATTTTGAGATTTT
CTGAAATTTACGCAAAAAATTTTCATTTTTTCAGCCACCTTCATCACTCTCGA
ATGATCGATCTCTTCACGTCAAATGCACTTTTTTCTGGATTTTTTTGTAAAAA

FIGURE 27

ATTTGAAATTCTCGTGTTTTTCTTCTGAAAAATTGCTTTTTTTGATTTTTTCTG
TAATTTTTTTTTTGTGATTTTCTTAATTTTTTAAATTTTCAAAAAATCTTTTC
ATCTCTTTCTCTCTCTCTCTGAATCTCAATTTTTTCTGAATTTCCCCGTTTTT
TCTGATAATTTCAATATTTCTCTGAATTTTCTATTCCCCCGTTGTAATGCC
AAAATATGTGGTAATTTCTCCCATTTTTTCGCTTTATTACTATTTATTCTATT
CAATTGGTGCCTCTCTCAATGTGTTGTATGAAAAACACTGTTTTATGGAGGTT
TTGGAGAATTTTGAATTTTTTCGTCTGTGATTTTTATTGGTTTTCTTTACCAATT
CAATTTTTTTTTTAATTCGAAAATTTGTAGAAATTCACTTTTGTAGCTTAAAAA
ATTAAAAATTGAGAAAATTTGTTCAAAAATGGCAAAGTTTTCGAAATTTTAGT
CTAAAAAAAGATTTTTTTAATATAGAATTTTAAAAAATTAGCACAGAAAAAT
GCCGAAAAATTCGTAATTTTTTCATTTAAAAATGAAAAAAAAAAAAACAAAA
AAAAAAAAAAAAAGAGGGAAAAATCCCATTTAAAAGTAGTTTTTTGACTGC
AAAATCGTCTGGAAATTAACAAAATTTAAAAAAATCTTTTTTACAGCCCATCG
TTTCCAAAAACCAAATAAAATGCCAAAAAAAAAATTTTTATGCAAAAATTCTG
GATTTTTTTCCGATTTTTTCAAAAAATCCCCCTTCTAAAAAAATGGTGAAT
TTGTTCCCAAAAACCAAAAATTTGAGATTTTCTAAAATTTTGGCAAAAATTAA
GAATTTACGGTTTTTGAGAGGGAAAACTCCATTAAAATTGATGATTTTATGA
CTAAAAATTCCTAAAAAATCAATTTTCAGTCAAAAATTAAATTT

Figure 28

MSEVIDESILNTEASDDPIPLNDDQIAELLGEDGEIMEITEQKDESDDVVMLDDD
DDDTPEPILVIDMDEDEDVTTDGPESQEELAADAPAPGAPEASAPAEASEASAP
DQEAPEVQDVPDSSGAPDASQAASEASDASAPEVPGSTEAQDAQDVPDSLGLASD
ASAEIPEAPEAPEAPEIAAEIDEEVLLAEQNGVLDEGFDETDDIIIEEAVEEAEA
VEPPINTENQENALEMLEERLKKNEEKEIVEKSDVKPEDEDIHHMETDSVETSSRK
RTGGATSPRSPAQKRPKRRVQTLLKMRQNAIELLTRLYGSWDAQLSLSNLETIRL
LGVNNNRKLIIEFEENEQVLKQKVSALTEELKKEKLAHAGTRSALKELTNEITGM
RVQMKNLRSMTQPTTSKIIDSFVQRHQAFEQQQQFQHQHHQHRPIMLAPRHHHP
PPPHFTPNQRAAAPYHPNMVQPNRLAAMPHRRPIIGMQQQNSAPPQFNHQAAL
VPSPQSSSAFSRPPPTQLATQRRAPPLASTGLPATVRWEAIPPKNPNVGHNEPPL
NNGGRAQPLIDNTRVHDNTIMLCVPLVSTANTISSGDSTRLPKVPRIYENLTANPD
LSVTIHSSAQDFRENYQIGGKINYEYLGGFQYNIQVFVQVSSLKFTGMNGYPDP
EDRISIDWGCSKLWPCKPKSHHKFRVRFHQAQLLPKNDRITIVAVAKDKTSGIHHI
SQPTFITLE

Figure 29

```

1  aaggaattag actctttatc taaagtgaag aatgatcaat taagaagttt ttgtcccata
61  gaattaaata taaatggatc tcctggggca gaatctgatt tggcaacatt ttgcacttct
121 aaaactgatg ctgttttaat gacttctgat gatagtgtga ctggatcgga attatcccct
181 ttgggtcaaaag catgcatgct ttcacaaat ggatttcaga atattagtag gtgcaaagaa
241 aaagacttgg atgataacctg catgctgcat aagaagtcag aaagcccatt tagagaaaca
301 gaacctctgg tgtcaccaca ccaagataaa ctcatgtcta tgccagttat gactgtggat
361 tattccaaaa cagtagttaa agaaccagtt gatacgaggg tttcttgctg caaaaccaa
421 gattcagaca tatactgtac tttgaacgat agcaaccctt ctttgtgtaa ctctgaagct
481 gaaaatattg agccttcagt tatgaagatt tcttcaaata gctttatgaa tgtgcatttg
541 gaatcaaaac cagtatatg tgatagtaga aatttgacag atcactcaaa atttgcattg
601 gaagaatata agcagagcat cggtagcact agttcagctt ctgttaatca ttttgatgat
661 ttatatcaac ctattgggag ttcaggtatt gcttcatctc ttcagagctc tccaccagga
721 ataaaggtgg acagtctaac tctcttgaaa tgcggagaga acacatctcc agttctggat
781 gcagtgttaa agagtaaaaa aagttcagag tttttaaagc atgcagggaa agaaacaata
841 gtagaagtga gtagtgcact tcctgattca ggaaagggat ttgcttccag ggagaacagg
901 cgtaataatg ggttatctgg gaaatgtttg caagaggctc aagaagaagg gaattccata
961 ttgcctgaaa gaagaggaag accagaaatc tctttagatg aaagaggaga aggaggacat
1021 gtgcatactt ctgatgactc agaagttgta tttcttctt gtgatttgaa ttaaccatg
1081 gaagacagtg atgggtgaac ttatgcatta aagtgtgaca gtagtggtca tgccccagaa
1141 attgtgtcta cagttcatga agattattct ggctcttctg aaagttcaaa tgatgaaagt
1201 gattcagaag atacagattc ggatgatagc agtattccaa gaaaccgtct ccagtcgtgt
1261 gtgggtgtgc caaagaattc tactttgccc atggaagaaa caagtccttg ttcttctcgg
1321 agcagtcaaa gttatagaca ctattctgac cattgggaag atgagagatt ggagtcaggg
1381 agacatttgt atgaggaaaa atttgaaagt atagcaagta aagcctgtcc tcaaaactgat
1441 aagtttttcc ttcataaagg aacagagaag aatccggaaa tttcttttac acagtcagat
1501 agaaaacaaa tagataaccg cctgcctgaa ctttctcatc ctccagagtga tggggttgat
1561 agtacaagtc atacagatgt gaaatctgac cctctgggtc acccaaattc agaggaaacc
1621 gtgaaagcca aaataacctc taggcagcaa gaagagctgc caatttatte ttctgatttt
1681 gaagatgtcc caaataagtc ttggcaacag accactttcc aaaacaggcc agatagtaga
1741 ctgggaaaaa cagaattgag tttttcttcc tcttgtgaga taccacatgt ggatggcttg
1801 cactcatcag aagagctcag aaacttaggt tgggacttct ctcaagaaaa gccttctacc
1861 acgtatcagc aacctgacag tagctatgga gcttgtggtg gacacaagta tcagcaaaat
1921 gcagaacagt atgggtggac acgtgattac tggcaaggca atggttactg ggatccaaga
1981 tcaggtagac ctcttggaa cgggggtgtg tatgatcgaa ctcaaggaca agtaccagat
2041 tccctaacag atgactgtga agaagaggag aattgggata aacaggatgg atcccatttt
2101 tcagaccagt ccgataaatt tcttctatcc cttcagaaa acaaggggtc agtgcaagca
2161 cctgaaataa gcagcaattc cattaaggac actttagctg tgaatgaaaa gaaagatttt
2221 tcaaaaaact tagaaaaaaa tgatatcaaa gatagagggc ctcttaaaaa aaggaggcag
2281 gaaatagaga gtgattctga aagtgatggt gagcttcagg acagaaagaa agttagagtg
2341 gaggtagagc agggagagac atcagtgcac ccagggtcag cactggttgg gccctcctgt
2401 gtcattggatg acttcaggga cccacagcga tgggaaggaat gtgccaagca agggaaaatg
2461 ccatgttact ttgatcttat tgaagaaaat gtttatttaa cagaaagaaa gaagaataaa
2521 tctcatcgag atattaagcg aatgcagtgt gagtgtacac ctctttctaa agatgaaaga
2581 gctcaaggtg aaatagcatg tggggaagat tgtcttaatc gtcttctcat gattgaatgt
2641 tcttctcggg gtccaaatgg ggattattgt tccaatagac ggtttcagag aaaacagcat
2701 gcagatgtgg aagtcatact cacagaaaag aaaggctggg gcttgagagc tgccaaagac
2761 cttccttcga acacctttgt cctagaatat tgtggagagg tactcgatca taaagagttt
2821 aaagctcgag tgaaggagta tgcacgaaac aaaaacatcc attactattt catggccctg
2881 aagaatgatg agataataga tgccactcaa aaaggaaatt gctctcggtt catgaatcac
2941 agctgtgaac caaattgtga aacccaaaaa tggactgtga acggacaact gagggttggg
3001 tttttacca ccaaactggt tccttcaggg tcagagttaa cgtttgacta tcagttccag
3061 agatatggaa aagaagccca gaaatgtttc tgcggatcag ccaattgccg gggttacctg
3121 ggaggagaaa acagagtcag catcagagca gcaggaggga aaatgaagaa ggaacgatct
3181 cgtaagaagg attcagtgga tggagagcta gaagctctga tggaaaatgg tgaggggtctc
3241 tctgataaaa accagggtgct cagcttatcc cggctaattg ttagaattga aactttggag

```

FIGURE 29

```

3301 cagaaactta cctgtctgga actcatcacag aacacacact cacagtcctg cctgaagtcc
3361 tttctggaac gtcattgggct gtctttgttg tggatctgga tggcagagct aggtgacggc
3421 cgggaaagta accagaagct tcaggaagag attataaaga ctttggaaaca cttgcccatt
3481 cctactaaaa atatgttgga ggaaagcaaa gtacttccaa ttattcaacg ctggtctcag
3541 actaagactg ctgtccctcc gttgagtga ggagatgggt attctagtga gaatacatcg
3601 cgtgctcata caccactcaa cacacctgat ccttccacca agctgagcac agaagctgac
3661 acagacactc ccaagaaaact aatgtttcgc agactgaaaa ttataagtga aaatagcatg
3721 gacagtgcaa tctctgatgc aaccagtga ctagaaggca aggatggcaa agaggatctt
3781 gatcaattag aaaatgtccc tgtagaggaa gaggaagaat tgcagtcaca acagctactc
3841 ccacaacagc tgcctgaatg caaagttgat agtgaaacca acatagaagc tagtaagcta
3901 cctacatctg aaccagaagc tgacgctgaa atagagctca aagagagcaa cggcacaaaa
3961 ctagaagaac ctattaatga agaaacacca tcccaagatg aagaggaggg tgtgtctgat
4021 gtggagagtg aaaggagcca agaacagcca gataaaacag tggatataag tgatttggcc
4081 accaaaactc tggacagttg gaaagaccta agggaggtat atcgaattcc aaagaaaagt
4141 caaactgaaa aggaaaacac aacaactgaa cgaggaaggg atgctgttgg cttcagagat
4201 caaacactg cccgaagac tcctaatagg tcaagagaga gagaccaga caagaaaact
4261 caaaataaag agaaaaggaa acgaagaagc tccctctcac caccctcttc tgcctatgag
4321 cggggaacaa aaaggccaga tgacagatat gatacaccaa cttctaaaaa gaaagtacga
4381 attaaagacc gcaataaact ttctacagag gaacgccgga agttgtttga gcaagaggtg
4441 gctcaacggg aggtctcaga acaacagcaa cagatgcaga acctgggaat gacatcacca
4501 ctgccctatg actctcttgg ttataatgcc ccgcatcatc cctttgctgg ttaccacca
4561 ggttatccca tgcaggccta tgtggatccc agcaacccta atgctggaaa ggtgctcctg
4621 cccacacca gcatggaccc agtgtgttct cctgctcctt atgatcatgc tcagcccttg
4681 gtgggacatt ctacagaacc cctttctgcc cctccaccag taccagtggg gccacatgtg
4741 gcagctcctg tggaaagttc cagttcccag tatgtggccc agagtgatgg tgtagtacac
4801 caagactcca gcgttgctgt cttgccagtg ccggcccccg gccagttca gggacagaat
4861 tatagtgttt gggattcaaa ccaacagtct gtacgtgtac agcagcagta ctctcctgca
4921 cagtctcaag caacctatata ttatcaagga cagacatgtc caacagtcta tgggtgtgaca
4981 tcaccttatt cacagacaac tccaccaatt gtacagagtt atgccagcc aagtcttcag
5041 tatatccagg ggcaacagat ttccacagct catccacaag gagtgggtgg acagccagcc
5101 gcagcagtga ctacaatagt tgcaccaggg cagcctcagc ccttgccagc atctgaaatg
5161 gttgtgacaa ataactctct ggatctgccc cccccccttc ctcccaaacc aaaaaccatt
5221 gtcttacctc ccaactggaa gacagctcga gatccagaag ggaagattta ttactaccat
5281 gtgatcacaa ggagactca gtgggacct cctacttggg aaagcccagg agatgatgcc
5341 agccttgagc atgaagctga gatggacgt ggaactccaa catatgatga aaacccatg
5401 aaggcctcga aaaagcccaa gacagacaa cgagacacct ccagtgaact agcaaagaaa
5461 agcaaagaag tattcagaaa agagatgtcc cagttcatcg tccagtgcct gaacccttac
5521 cggaaacctg actgcaaagt gggaaagaatt accacaactg aagactttaa acatctggct
5581 cgcaagctga ctcacggtgt tatgaataag gagctgaagt actgtaagaa tctgaggac
5641 ctggagtgc atgagaatgt gaaacacaaa accaaggagt acattaagaa gtacatgcag
5701 aagtttgggg ctgtttacaa acccaaagag gacactgaat tagagtgact gttgggcccag
5761 ggtgggagga tgggtggtca ggtaagacag actctaggga gaggaaatcc tgtgggcctt
5821 tctgtccac ccctgtcagc actgtgctac tgatgataca tcaccctggg gaattcaacc
5881 ctgcagatgt caactgaagg ccacaaaaat gaactccatc tacaagtgat tacctagttg
5941 tgagctgttg gcatgtggtt agaagccatc agaggtgcaa gggcttagaa aagaccctgg
6001 ccagacctga ctccactctt aaacctgggt cttctccttg gcggtgctgt cagcgcacag
6061 acccatgcgc atccccacce acaacctttt accctgatga tctgtattat attttaatgt
6121 atatgtgaat atattgaaaa taatttgttt tttcctgggt tttgtttggg tttcgttttg
6181 cttttagcct ctacatgcta ggatcacagg aagactttgt aaggacagtt taagttctcc
6241 tgcaagggtt aatttgttat catgtaataa ttccaaagca ggctgccttg tgggttttggc
6301 cagccttggt ctatgttgat aagattgatt tactgcttaa aatcacttta ctttatccaa
6361 tttttactga actttttatg taaaaaata aaatcaatta aag

```

Figure 30

KELDSLKVKNDQLRSFCPIELNINGSPGAESDLATFCTSKTDAVLMTSDDSVTGSELSPLVKACMLSSNG
FQNI SRCKEKLDDTCMLHKKSESPFRETEPLVSPHQDKLMSMPVMTVDYSKTVVKEPVDTRVSCCKTKDS
DIYCTLNDSNPSLCNSEAENIEPSVMKISSNSFMNVHLESKPVICDSRNLTDHSKFACEEYKQSIGSTSSA
SVNHFDLLYQPIGSSGIASSLQSLPPGIKVDSLTLKCGENTSPVLDAVLKSKKSSEFLKHAGKETIVEVG
SDLPDSGKGFASRENRRNNGLSGKCLQEAQEEGNSILPERRGRPEISLDERGEGGHVHTSDDSEVVFFSSCD
LNLTMEDSDGVTYALKCDSSGHAPEIVSTVHEDYSGSSSESSNDESDSEDTSDDSSI PRNRLQSVVVVPKN
STLPMEETSPCSSRSSQSYRHYS DHWEDERLESRRHLYEEKFESIASKACPQTDKFFLHKGTEKNPEISFT
QSSRKQIDNRLPELSHPQSDGVDSTSHTDVKS DPLGHPNSEETVKAKIPSRQOEELPIYSSDFEDVPNKSW
QQTTFQNRPD SRLGKTELSFSSSCEIPHVDGLHSSEELRNLGWDFSQEK PSTTYQQPDSSYGACGGHKYQQ
NAEQYGGTRDYWQNGYWDPRSGRPPGTGVVYDRTOGQVPDSLTDREEEENWDQDGSFSDQSDKFLLS
LQDKGSGVQAPEISSNSIKDTLAVNEKKDFSKNLEKNDIKDRGPLKKRRQEI ESDES DGELQDRKKVRVE
VEQGETSVPPGSALVG PSCVMDDFRDPQRWKECAKQGMPCYFDLIEENVYLTERKKNKSHRDIKRMQCEC
TPLSKDERAQGEIACGEDCLNRLLMIECSSRCPNGDYCSNRRFQRKQHADVEVILTEKKGWGLRAAKDLPS
NTFVLEYCGEVL DHKEFKARVKEYARNKNIHYFFMALKNDEIIDATQKGNCSR FMNHSCEPNCETQKWTVN
GQLRVGFFTTKLVP SGSELTFDYQFQRYGKEAQKCFGSANCRGYLGGENRVSI RAAGGKMKKERSRKKDS
VDGELEALMENEGELSDKNQVLSLSRLMVRIETLEQKLTCELELIQNTHSQSCLKSFLERHGLSLLWIWMAE
LGDGRESNQKLQEEIIKTLEHLPIPTKNMLEESKVLPIIQRWSQTKTAVPPLSEGDGYSSENTSRAHTPLN
TPDPSTKLSTEADTDTPKKLMFRRLKIISENSMDSAISDATSELEGKDGKEDLDQLENVPVEEEEEELOSQO
LLPQQLPECKVDSETNIEASKLPTSEPEADAEIELKESNGTKLEEPINEETPSQDEEEGVSDVESERSQEQ
PDKTVDISDLATKL LDSWKDLKEVYRIPKKSQTEKENTTTERGRDAVGFRDQTPAPKTPNRSRERDPDKQT
QNKEKRKRSSSLSPSSAYERGTRPDDRYDTPTS KKKVRIKDRNKLSTEERRKLFEQEV AQREAAQKQQQQ
MQNLGMTSPLPYDSL GYNAPHHPFAGYPPGYPMQAYVDPSNPNAGKVLLPTPSMDPVCSPAPYDHAQPLVG
HSTEPLSAPPPVPVPHVAAPVEVSSSQYVAQSDGVVHQDSSVAVLPVPAPGPVQGNYSVWDSNQSVSV
QQQYSPAQSQATIYYQGQTCPTVYGVTS PYSQTTPIVQSYAQPSLQYIQGQQIFTAHPQGVVVPAAAVT
TIVAPGQPQPLQ PSEMVTNNLLDLPPPSPPKPKTIVLPPNWKTARDPEGKIYYYHVITRQTQWDPPTWES
PGDDASLEHEAEMDLGTPTYDENPMKASKKPKTAEADTSSELAKKSKEVFRKEMSQFIVQCLNPHYRKPDCK
VGRITTTEDFKHLARKLTHGVMNKELKYCKNPEDLECNENVKHKTKKEYIKKYMQKFGAVYKPKEDTELE

Confidently predicted domains, repeats, motifs and features:

name	begin	end	E-value
<u>Pfam:AT hook</u>	47	60	1.80E+01
<u>low complexity</u>	230	243	-
<u>low complexity</u>	327	338	-
<u>low complexity</u>	371	400	-
<u>low complexity</u>	505	530	-
<u>coiled coil</u>	549	621	-
<u>AWS</u>	636	682	8.80E-18
<u>SET</u>	683	811	6.00E-41
<u>PostSET</u>	812	828	7.40E-04
<u>low complexity</u>	1080	1093	-
<u>low complexity</u>	1118	1129	-
<u>low complexity</u>	1138	1158	-
<u>low complexity</u>	1271	1287	-
<u>VWV</u>	1361	1393	4.10E-08
<u>low complexity</u>	1447	1468	-
<u>low complexity</u>	1469	1497	-

These features and domains are not shown in the diagram, either because their scores are less significant than the required threshold, or because they overlap with some other source of annotation:

name	begin	end	E-value	reason
<u>low complexity</u>	36	50	-	overlap
<u>low complexity</u>	532	554	-	overlap
<u>low complexity</u>	569	615	-	overlap
<u>Pfam:SET</u>	677	811	8.80E-48	overlap
<u>low complexity</u>	734	739	-	overlap
<u>Pfam:VWV</u>	1362	1391	1.90E-08	overlap

Figure 31 LIN(n3628) Functional domains

Confidently predicted domains, repeats, motifs and features:

name	begin	end	E-value
<u>low complexity</u>	387	411	-
<u>low complexity</u>	435	449	-
<u>AWS</u>	845	900	7.50E-30
<u>SET</u>	901	1024	3.10E-41
<u>PostSET</u>	1025	1041	2.50E-05
<u>low complexity</u>	1262	1286	-
<u>low complexity</u>	1333	1344	-
<u>low complexity</u>	1425	1437	-
<u>coiled coil</u>	1468	1491	-
<u>low complexity</u>	1569	1589	-
<u>low complexity</u>	1605	1619	-
<u>low complexity</u>	1622	1643	-
<u>low complexity</u>	1690	1710	-
<u>WW</u>	1741	1773	2.10E-11

These features and domains are not shown in the diagram, either because their scores are less significant than the required threshold, or because they overlap with some other source of annotation:

name	begin	end	E-value	reason
<u>Pfam:SET</u>	895	1024	6.30E-52	overlap
<u>low complexity</u>	1477	1493	-	overlap
<u>low complexity</u>	1726	1744	-	overlap
<u>Pfam:WW</u>	1742	1771	6.90E-12	overlap

Figure 32 KIAA1732 Domains

SEQUENCE LISTING

<110> MASSACHUSETTS INSTITUTE OF TECHNOLOGY et al.

<120> RB PATHWAY AND CHROMATIN REMODELING
GENES THAT ANTAGONIZE LET-60 RAS SIGNALING

<130> 01997/548WO3

<150> 60/437,821

<151> 2003-01-02

<150> 60/410,160

<151> 2002-09-12

<160> 36

<170> FastSEQ for Windows Version 4.0

<210> 1

<211> 853

<212> PRT

<213> Caenorhabditis elegans

<400> 1

Met	Val	Thr	Ala	Asp	Glu	Thr	Val	Leu	Ala	Thr	Thr	Thr	Asn	Thr	Thr
1				5					10					15	
Ser	Met	Ser	Val	Glu	Pro	Thr	Asp	Pro	Arg	Ser	Ala	Gly	Glu	Ser	Ser
			20					25					30		
Ser	Asp	Ser	Glu	Pro	Asp	Thr	Ile	Glu	Gln	Leu	Lys	Ala	Glu	Gln	Arg
			35				40					45			
Glu	Val	Met	Ala	Asp	Ala	Ala	Asn	Gly	Ser	Glu	Val	Asn	Gly	Asn	Gln
	50					55					60				
Glu	Asn	Gly	Lys	Glu	Glu	Ala	Ala	Ser	Ala	Asp	Val	Glu	Val	Ile	Glu
65					70					75				80	
Ile	Asp	Asp	Thr	Glu	Glu	Ser	Thr	Asp	Pro	Ser	Pro	Asp	Gly	Ser	Asp
				85				90					95		
Glu	Asn	Gly	Asp	Ala	Ala	Ser	Thr	Ser	Val	Pro	Ile	Glu	Glu	Glu	Ala
			100					105					110		
Arg	Lys	Lys	Asp	Glu	Gly	Ala	Ser	Glu	Val	Thr	Val	Ala	Ser	Ser	Glu
			115				120					125			
Ile	Glu	Gln	Asp	Asp	Asp	Gly	Asp	Val	Met	Glu	Ile	Thr	Glu	Glu	Pro
	130					135					140				
Asn	Gly	Lys	Ser	Glu	Asp	Thr	Ala	Asn	Gly	Thr	Val	Thr	Glu	Glu	Val
145					150					155				160	
Leu	Asp	Glu	Glu	Glu	Pro	Glu	Pro	Ser	Val	Asn	Gly	Thr	Thr	Glu	Ile
				165				170						175	
Ala	Thr	Glu	Lys	Glu	Pro	Glu	Asp	Ser	Ser	Met	Pro	Val	Glu	Gln	Asn
			180					185					190		
Gly	Lys	Gly	Val	Lys	Arg	Pro	Val	Glu	Cys	Ile	Glu	Leu	Asp	Asp	Asp
			195				200					205			
Asp	Asp	Asp	Glu	Ile	Gln	Glu	Ile	Ser	Thr	Pro	Ala	Pro	Ala	Lys	Lys
	210					215					220				
Ala	Lys	Ile	Asp	Asp	Val	Lys	Ala	Thr	Ser	Val	Pro	Glu	Glu	Asp	Asn
225					230					235				240	
Asn	Glu	Gln	Ala	Gln	Lys	Arg	Leu	Leu	Asp	Lys	Leu	Glu	Glu	Tyr	Val

				245					250				255				
Lys	Glu	Gln	Lys	Asp	Gln	Pro	Ser	Ser	Lys	Ser	Arg	Lys	Val	Leu	Asp		
			260					265					270				
Thr	Leu	Leu	Gly	Ala	Ile	Asn	Ala	Gln	Val	Gln	Lys	Glu	Pro	Leu	Ser		
		275					280					285					
Val	Arg	Lys	Leu	Ile	Leu	Asp	Lys	Val	Leu	Val	Leu	Pro	Asn	Thr	Ile		
	290					295					300						
Ser	Phe	Pro	Pro	Ser	Gln	Val	Cys	Asp	Leu	Leu	Ile	Glu	His	Asp	Pro		
305					310					315					320		
Glu	Met	Pro	Leu	Thr	Lys	Val	Ile	Asn	Arg	Met	Phe	Gly	Glu	Glu	Arg		
			325					330						335			
Pro	Lys	Leu	Ser	Asp	Ser	Glu	Lys	Arg	Glu	Arg	Ala	Gln	Leu	Lys	Gln		
			340					345					350				
His	Asn	Pro	Val	Pro	Asn	Met	Thr	Lys	Leu	Leu	Val	Asp	Ile	Gly	Gln		
		355					360					365					
Asp	Leu	Val	Gln	Glu	Ala	Thr	Tyr	Cys	Asp	Ile	Val	His	Ala	Lys	Asn		
	370					375					380						
Leu	Pro	Glu	Val	Pro	Lys	Asn	Leu	Glu	Thr	Tyr	Lys	Gln	Val	Ala	Ala		
385					390					395					400		
Gln	Leu	Lys	Pro	Val	Trp	Glu	Thr	Leu	Lys	Arg	Lys	Asn	Glu	Pro	Tyr		
			405					410						415			
Lys	Leu	Lys	Met	His	Arg	Cys	Asp	Val	Cys	Gly	Phe	Gln	Thr	Glu	Ser		
			420					425					430				
Lys	Leu	Val	Met	Ser	Thr	His	Lys	Glu	Asn	Leu	His	Phe	Thr	Gly	Ser		
		435					440					445					
Lys	Phe	Gln	Cys	Thr	Met	Cys	Lys	Glu	Thr	Asp	Thr	Ser	Glu	Gln	Arg		
	450					455					460						
Met	Lys	Asp	His	Tyr	Phe	Glu	Thr	His	Leu	Val	Ile	Ala	Lys	Ser	Glu		
465					470					475					480		
Glu	Lys	Glu	Ser	Lys	Tyr	Pro	Cys	Ala	Ile	Cys	Glu	Glu	Asp	Phe	Asn		
			485					490						495			
Phe	Lys	Gly	Val	Arg	Glu	Gln	His	Tyr	Lys	Gln	Cys	Lys	Lys	Asp	Tyr		
			500					505						510			
Ile	Arg	Ile	Arg	Asn	Ile	Met	Met	Pro	Lys	Gln	Asp	Asp	His	Leu	Tyr		
		515					520					525					
Ile	Asn	Arg	Trp	Leu	Trp	Glu	Arg	Pro	Gln	Leu	Asp	Pro	Ser	Ile	Leu		
	530					535					540						
Gln	Gln	Gln	Gln	Gln	Ala	Ala	Leu	Gln	Gln	Ala	Gln	Gln	Lys	Lys	Gln		
545					550					555					560		
Gln	Gln	Leu	Leu	His	Gln	Gln	Gln	Ala	Ala	Gln	Ala	Ala	Ala	Ala	Ala		
			565					570						575			
Gln	Leu	Leu	Arg	Lys	Gln	Gln	Leu	Gln	Gln	Gln	Gln	Gln	Gln	Gln	Gln		
			580					585						590			
Ala	Arg	Leu	Arg	Glu	Gln	Gln	Gln	Ala	Ala	Gln	Phe	Arg	Gln	Val	Ala		
		595					600					605					
Gln	Leu	Gln	Gln	Gln	Gln	Ser	Ala	Gln	Ala	Gln	Arg	Ala	Gln	Gln	Asn		
	610					615					620						
Gln	Gly	Asn	Val	Asn	His	Asn	Thr	Leu	Ile	Ala	Ala	Met	Gln	Ala	Ser		
625					630					635					640		
Leu	Arg	Arg	Gly	Gly	Gln	Gln	Gly	Asn	Ser	Leu	Ala	Val	Ser	Gln	Leu		
			645					650						655			
Leu	Gln	Lys	Gln	Met	Ala	Ala	Leu	Lys	Ser	Gln	Gln	Gly	Ala	Gln	Gln		
		660						665						670			
Leu	Gln	Ala	Ala	Val	Asn	Ser	Met	Arg	Ser	Gln	Asn	Ser	Gln	Lys	Thr		
		675					680							685			
Pro	Thr	His	Arg	Thr	Pro	Thr	Phe	Val	Cys	Glu	Ile	Cys	Asp	Ala	Ser		
	690					695					700						
Val	Gln	Glu	Lys	Glu	Lys	Tyr	Leu	Gln	His	Leu	Gln	Thr	Thr	His	Lys		

705		710		715		720
Gln Met Val Gly	Lys Val Leu Gln Asp	Met Ser Gln Gly Ala	Pro Leu			
	725	730	735			
Ala Cys Ser Arg	Cys Arg Asp Arg Phe Trp Thr Tyr Glu Gly	Leu Glu				
	740	745	750			
Arg His Leu Val	Met Ser His Gly Leu Val Thr Ala Asp	Leu Leu Leu				
	755	760	765			
Lys Ala Gln Lys	Lys Glu Asp Gly Arg Cys Lys Thr Cys Gly Lys					
	770	775	780			
Asn Tyr Ala Phe	Asn Met Leu Gln His Leu Val Ala Asp His Gln Val					
	785	790	795			
Lys Leu Cys Ser	Ala Glu Ile Met Tyr Ser Cys Asp Val Cys Ala Phe					
	805	810	815			
Lys Cys Ser Ser	Tyr Gln Thr Leu Glu Ala His Leu Thr Ser Asn His					
	820	825	830			
Pro Lys Gly Asp	Lys Lys Thr Ser Thr Pro Ala Lys Lys Asp Asp Cys					
	835	840	845			
Ile Thr Leu Asp Asp						
	850					

<210> 2

<211> 4001

<212> DNA

<213> Caenorhabditis elegans

<400> 2

tcacacactc	atgacataca	cacatcattt	cgccctcacac	accgcgcgct	cgccatccgc	60
accgcccggg	tgggacgtgt	tcaaactttt	cggttttcgt	aattaatagt	gagccccggg	120
ttattcgctt	tgagaatcag	tataatggat	atatcagatt	gtgtaattag	gttgcggtgct	180
tgaactttta	aaattaactg	ttttaaat	atctgccttt	atcgttacag	taaatcattt	240
tgatgaactt	ttcggatgaa	tcataatgaa	gtacgcagcg	ctctaacaaa	atgtgtttgt	300
aaattccaat	tgctacaagt	tgcccggctt	atTTTTTggg	gattgaagca	tgattctgtt	360
gacgctcccc	acgcggaata	ccaggacgga	ccgatgagag	agtactgcca	gtgaagagac	420
gcatgcgagc	aggacgagtg	ctgctcaccc	ttcttctcag	cgtcggcgcc	tgcgaccagc	480
ggccgaggaa	ggggaggaga	gaggccgatt	tggtgcgcta	ccacgtttga	tactcagtca	540
cttaccacag	ctggttctct	tgtgcgttca	aatctggctt	gccgcgcgcg	cgcattttat	600
tcctaccagt	ttgaatctcc	cacctctccg	actgtaactg	tcctaatttg	cttccttctc	660
atcaactctct	ctttgcctat	ttctcactat	ctagactcta	tttttccaga	atggtcaccg	720
ccgacgagac	ggtactcgcc	acaacgacca	acaccacttc	catgtctgtg	gaaccaacgg	780
atccgagaag	cgctggtgaa	tcgtcctcag	attcggagcc	agacacaatt	gaggtgagga	840
aaagttttgg	gaattttaa	ctgaataaaa	cgttttcagc	agctgaaggc	agaacagcgc	900
gaagtgatgg	ccgacgcggc	gaatggttcc	gaagtcaacg	gaaatcaaga	gaacggaaaa	960
gaggaagcgg	catctgcaga	cgtggaagtg	atcgagatag	atgacaccga	agagtctacg	1020
gatccctcac	ctgatggatc	tgatgaaaac	ggtgatgctg	catctacatc	ggttccaatc	1080
gaagaggaag	cgcgtaaaaa	ggatgagggg	gcttccgaag	tgactgtggc	atcatctgag	1140
attgaacaag	acgatgatgg	cgatgttatg	gaaatcactg	aggagccgaa	cggaaaagtcg	1200
gaggatactg	ccaacggaac	aggtgtgttt	tataatttta	ccaagttaa	ttttaacttt	1260
ctattttccag	ttactgagga	ggtgctagat	gaagaggagc	cagaaccttc	cgtaaacgga	1320
acaactgaga	tcgctacaga	gaaagagcca	gaagattctt	caatgcctgt	cgaacagaat	1380
gggaaggggtg	tgaagcggcc	tgtcgaatgc	atcgaactcg	acgacgacga	tgatgacgag	1440
attcaggaaa	tttctacccc	tgccccagct	aaaaaagcta	aaattgatga	tgtcaaggcg	1500
acaagcggttc	cagaagagga	caacaatgag	caggcgcaga	agagattgct	cgacaagctg	1560
gaagagtatg	tgaaggagca	gaaggatcaa	ccatccagca	aaagccgaaa	agttctggag	1620
actcttctcg	gagcaatcaa	tgcgcaagtt	caaaaggagc	ctctgtcggt	tcggaagctg	1680
atcctggaca	aagttctcgt	tctcccaaac	acaatatcat	tcccaccaag	tcaagtttgc	1740
gacttattga	ttgagcacga	tcccgaatg	cctttgacga	aggttatcaa	caggatgttt	1800
ggagaagaaa	gaccaaagtt	gagtgattcc	gagaaacgag	agagagctca	gctgaaacaa	1860

cataatcctg	ttccaaatat	gacaaaactg	ctcgtggaca	ttggacagga	tctcgttcaa	1920
gaagctacct	attgtgatat	agttcacgcg	aagaatcttc	cagaggtgcc	aaaaaatctt	1980
gaaacctata	agcaagtgcg	tgcgcagttg	aaaccagttt	gggagacatt	gaaacgcaaa	2040
aatgagccgt	acaagttgaa	aatgcatcga	tgcgacgtct	gtggattcca	gacggaatca	2100
aagctgggta	tgagcactca	caaggagaat	ttgcacttca	caggatccaa	attccagtg	2160
accatgtgta	aagagacgga	cacgagtgag	caaagaatga	aggatcacta	cttgttaagt	2220
tttttttttt	catctttcaa	tattcattta	attacagcga	aactcatctt	gttattgcaa	2280
aatcggaaga	gaaggagtcc	aagtatccat	gtgcaatctg	cgaagaagac	ttcaatttca	2340
aaggtgtccg	tgagcagcat	tacaagcagt	gcaagaagga	ctacattcgc	attcgaaaca	2400
tcattgatgc	gaagcaagac	gatcatctct	atatcaacag	atgggtctgg	gagagcccc	2460
aattggatcc	cagcattctt	caacagcagc	aacaagctgc	tcttcagcaa	gctcaacaaa	2520
agaagcaaca	gcaacttctg	catcaacagc	aagcagcaca	agctgcagcc	gctgcgcaac	2580
tcttacggaa	gcaacaatta	caacagcaac	aacaacagca	acaggctcgt	cttcgtgagc	2640
aacagcaagc	ggcccaattc	cggcaagtgg	ctcaactgct	gcaacaacaa	tcagcgcagg	2700
ctcaacgtgc	acagcagaat	caaggaaatg	tgaatcataa	cactctgatt	gcaggtaata	2760
gctaaacata	ttttaaataa	gtattttgtg	taattattta	tatttcagca	atgcaagcgt	2820
cgttgcgtag	aggtggtcaa	caaggaaatt	cgtggtcagt	ttctcaactt	ctccaaaagc	2880
aaatggcagc	tttgaagtgc	caacaaggag	ctcaacaact	tcaggctcgc	gtgaactcca	2940
tgagaagcca	gaacagtcaa	aagacgccaa	cacacagaag	ttcgaaactt	gttactacgc	3000
cgtctcatgc	tactgttggc	tcttcttcag	ctccacggtt	tgtatgcgaa	atttgtgatg	3060
cgtcagtgca	ggaaaaggag	aagtatctac	agcatcttca	ggtaatttta	agaaacgttt	3120
ctattttcaat	ttcaaaaaccg	attattaaat	atcttaacaa	tcacattttc	agactactca	3180
taagcagatg	gttggaaaag	tgtgcagga	catgtcgcaa	ggagctccac	tggcatgttc	3240
tcgatgccgt	gacagattct	ggacttatga	agggttggag	cggcacttgg	tgatgtcgca	3300
tggtctcgtc	actgctgatc	tgtctctcaa	agcgcaaaaag	aagggaagacg	gaggtcgatg	3360
caagacatgc	ggcaagaact	atgcgttcaa	catgcttcaa	cacttggtag	ctgatcatca	3420
agtgaagtgt	tgtcgggtg	aatcatgta	ctcgtgcgat	gtgtgcgcgt	tcaaatgttc	3480
gagttatcag	actctggaag	cccatctcac	ttcaaatcac	ccaaaaggag	ataagaagac	3540
atcaacacca	gcataaaaaag	atgattgtat	tactctggat	gattaatagg	aaaacgaatg	3600
gcttatcccg	ttctacgaat	gagtgttgga	aacattcttc	acaatgatct	caattatttc	3660
tcttattctt	tacattcaat	catttttaaat	caccagttct	cccactttca	ttgatataca	3720
cattctattg	cgggttccgg	aaccgaaatc	aatcagtact	ttactttatt	tccccaattt	3780
ttctcttcat	gatattctgg	ttattctcgc	atcttcccct	accttcaaaa	ctccctattt	3840
ttttttcaaa	acctaactac	cccacaatta	tcattgtaaaa	tcaaatgtga	attccccata	3900
agacagatca	gtatacactt	tcacttcata	cgtctgttgt	tctcccccat	ctcactactt	3960
ttttaccatt	tgtccagtta	agatttttgg	aagatatcta	t		4001

<210> 3

<211> 2562

<212> DNA

<213> Caenorhabditis elegans

<400> 3

atgggtcacg	ccgacgagac	ggtactcgcc	acaacgacca	acaccacttc	catgtctgtg	60
gaaccaacgg	atccgagaag	cgtggtgaa	togtctcag	attcggagcc	agacacaatt	120
gagcagctga	aggcagaaca	gcgcgaagt	atggccgacg	cggcgaatgg	ttccgaagtc	180
aacggaaatc	aagagaacgg	aaaagaggaa	gcggcatctg	cagacgtgga	agtgatcgag	240
atagatgaca	ccgaagagtc	tacggatccc	tcacctgatg	gatctgatga	aaacgggtgat	300
gctgcatcta	catcggttcc	aatcgaagag	gaagcgcgta	aaaaggatga	gggggcttcc	360
gaagtgactg	tggcatcatc	tgagattgaa	caagacgatg	atggcgatgt	tatggaaatc	420
actgaggagc	cgaacggaaa	gtcggaggat	actgccaacg	gaacagttac	tgaggagggtg	480
ctagatgaag	aggagccaga	accttccgta	aacggaaacaa	ctgagatcgc	tacagagaaa	540
gagccagaag	attcttcaat	gcctgtcgaa	cagaatggga	aggggtgtgaa	gcggcctgtc	600
gaatgcatcg	aactcgacga	cgacgatgat	gacgagattc	aggaaaatttc	taccctgtcc	660
ccagctaaaa	aagctaaaaat	tgatgatgtc	aaggcgacaa	gcgttccaga	agaggacaac	720
aatgagcagg	cgcagaagag	attgctcgac	aagctggaag	agtatgtgaa	ggagcagaag	780
gatcaaccat	ccagcaaaag	ccgaaaagtt	ctggacactc	ttctcggagc	aatcaatgcg	840
caagttcaaa	aggagcctct	gtcggttcgg	aagctgatcc	tggacaaaagt	tctcgttctc	900

```

ccaaacacaa tatcattccc accaagtcaa gtttgcgact tattgattga gcacgatccc 960
gaaatgcctt tgacgaaggt tatcaacagg atgtttggag aagaaagacc aaagttgagt 1020
gattccgaga aacgagagag agctcagctg aaacaacata atcctgttcc aaatatgaca 1080
aaactgctcg tggacattgg acaggatctc gttcaagaag ctacctattg tgatatagtt 1140
cacgcgaaga atcttccaga ggtgccaaaa aatcttgaaa cctataagca agtcgctgcg 1200
cagttgaaac cagtttggga gacattgaaa cgcaaaaatg agccgtacaa gttgaaaatg 1260
catcgatgcg acgtctgtgg attccagacg gaatcaaagc tggttatgag cactcacaag 1320
gagaatttgc acttcacagg atccaaattc cagtgcacca tgtgtaaaga gacggacacg 1380
agtgagcaaa gaatgaagga tcactacttc gaaactcatc ttgttattgc aaaatcggaa 1440
gagaaggagt ccaagtatcc atgtgcaatc tgcgaagaag acttcaattt caaagggtgc 1500
cgtgagcagc attacaagca gtgcaagaag gactacattc gcattcgaaa catcatgatg 1560
ccgaagcaag acgatcatct ctatatcaac agatggctct gggagaggcc ccaattggat 1620
cccagcattc ttcaacagca gcaacaagct gctcttcagc aagctcaaca aaagaagcaa 1680
cagcaacttc tgcataca gcaagcagca caagctgcag ccgctgcgca actcttacgg 1740
aagcaacaat tacaacagca acaacaacag caacaggctc gtcttcgtga gcaacagcaa 1800
gcggcccaat tccggcaagt ggctcaactg ctgcaacaac aatcagcgca ggctcaacgt 1860
gcacagcaga atcaaggaaa tgtgaatcat aacactctga ttgcagcaat gcaagcgctc 1920
ttgcgtagag gtggtcaaca aggaaattcg ctggcagttt ctcaacttct ccaaaagcaa 1980
atggcagctt tgaagtcgca acaaggagct caacaacttc aggctgcggt gaactccatg 2040
agaagccaga acagtcaaaa gacgccaaca cacagaactc ccacgtttgt atgcgaaatt 2100
tgtgatgcgt cagtgcagga aaaggagaag tatctacagc atcttcagac tactcataag 2160
cagatggttg gaaaagtgtc gcaggacatg tcgcaaggag ctccactggc atgttctcga 2220
tgccgtgaca gattctggac ttatgaaggg ttggagcggc acttggtgat gtgcgatgg 2280
ctcgtcactg ctgatctgct cctcaaagcg caaaagaagg aagacggagg tcgatgcaag 2340
acatgcggca agaactatgc gttcaacatg cttcaacact tggtagctga tcatcaagt 2400
aagttgtgct cggctgaaat catgtactcg tgcgatgtgt gcgcgttcaa atgctcgagt 2460
tatcagactc tggaagccca tctcacttca aatcacccaa aaggagataa gaagacatca 2520
acaccagcaa aaaaagatga ttgtattact ctggatgatt aa 2562

```

<210> 4

<211> 10

<212> DNA

<213> Caenorhabditis elegans

<400> 4

agtttcagac

10

<210> 5

<211> 10

<212> DNA

<213> Caenorhabditis elegans

<400> 5

agtttcagac

10

<210> 6

<211> 13

<212> DNA

<213> Caenorhabditis elegans

<400> 6

agaatcttca gtc

13

<210> 7

<211> 13

<212> DNA

<213> Caenorhabditis elegans

<400> 7
agaatcttca gcc 13

<210> 8
<211> 13
<212> DNA
<213> Caenorhabditis elegans

<400> 8
agaactttaa gat 13

<210> 9
<211> 13
<212> DNA
<213> Caenorhabditis elegans

<400> 9
agaactttaa gat 13

<210> 10
<211> 10
<212> DNA
<213> Caenorhabditis elegans

<400> 10
agttgcagaa 10

<210> 11
<211> 10
<212> DNA
<213> Caenorhabditis elegans

<400> 11
agttgcagaa 10

<210> 12
<211> 16061
<212> DNA
<213> Caenorhabditis elegans

<400> 12
gaggaagatg tagacgacga ttcgggtttcc gtactctcat gactttttggc gaaaatcctc 60
acgaattctt tttccgtcat acgttgagtt aaaaatctgg cgatgtaacg aagaatgaga 120
agagcgtttg atgtttgcca taagtagatt ttactgaaat aagaaaaagc ttttaattaaa 180
tataatgatg attttttttt ccaactcact tttcgcattg ttctgatgtt tttagttctg 240
tggctctgcg aaggaaaaagt cgaataaaatg cagcgaaatt tcctgttggt tgtgtattgt 300
acattagaca ttgaagatga tcatctaaag cagattccaa agcgattcgg gtgtctctaa 360
acgattataa cattttttaa gctttttgcct aattttaatc cttactcgtc gtcatcatca 420
aacttgagac tgaaagagag aagtttgttc caaaatgggt cataatcgtc gacaggttcc 480
aaaccgctga gtttcttcag ataaatatc tcctgtaaga ccgtttcctt gggtataact 540
gatcccatgt gtctgaaatt tgttattaca ctgtaataa tcataaaaaat aaaagaaaaa 600
gtcaagaaag ggtcaaataat taatcaggtc acatcttttt tattcaataa aatctcctct 660
ctcgttcgtg gcaatgcacg tgaaatgcgc caacaacögc gāgēgcēdca ācacacacac 720
atacgcgtca gcagacaatt cgctctcgtt tgaaatttag ttgtttcttt gtttctgctg 780
aaataatgtc agttttccga taatttcagc gttttctgac tgatttttct tgttgcattc 840
acttcctaata agttcattct actccattct tcattttata atctgtttcc ttcgcaattt 900
agtgaattaa acacgtaaat cttgtttcag ataaattatt caaatagttg caciaagctc 960
aatagtttag aagtatcttc agtgctggtc actaatataa aatggatccg gctatggctt 1020

ctccaggcta	tccgtctgtg	cagtccgatc	ggagtaatca	cctaacagag	ctggaaacga	1080
gaattcaaaa	tcttgccgat	aattcacaaa	gagatgatgt	caaattgaaa	atgttacaaag	1140
ttagtttcaa	taattcgtgt	taagtaatca	atttgttcgg	ttgcaggaga	tttggagcac	1200
aatcgaaaat	catttcacac	taagttcgca	cgagaaagtc	gtggagaggc	tcattctctc	1260
gttcctacaa	gttttctgca	acacaagtcc	acagttcatt	gctgaaaaca	atacacaaca	1320
gcttcgaaag	ttaatgcttg	aaatcattct	tgcactttcg	aacgtagaag	ccatgaaaca	1380
tcatagcaaa	gaaattatca	agcagatgat	gaggctaata	accgtggaaa	atgaggagaa	1440
tgccaatttg	gctatcaaaa	ttgtcaccga	tcaagggaga	agtaccggca	aaatgcaata	1500
ttgcggagag	gttttcacaga	taatggctct	cttcaaaaaca	atgggtcattg	atctgacggc	1560
gagtggtcga	gctgggtgata	tgttcaacat	aaaagagcat	aaagctccac	cgtcaactag	1620
ctccgacgag	caagtcacat	ctgaatatct	gaagacttgc	tactatcaac	aaacggttct	1680
tctcaacgga	acggaaggaa	aaccgccatt	aaaatacaat	atgattccat	cagctcatca	1740
gtcaacgaag	gtgctcctgg	aggttccgta	tctcgtgatt	ttcttctatc	aacatttcaa	1800
aacagcgatc	caaaccgaag	cgcttgattt	catgaggctt	ggtcttgatt	ttctaaatgt	1860
cagagttcca	gacgaggata	aactcaaaac	aatcaaaata	ataaccgatg	attttgtcag	1920
tgacagtcct	cgattcctgt	cattcgtcaa	cattatggct	aagattccag	cggtaagttt	1980
cgtttttttca	agtttttttt	ctgtaatcct	gattttttatt	tttcagttta	tggatcttat	2040
catgcaaaaat	ggaccgcttc	tagtgtcggg	aacaatgcag	atgctcgagc	ggtgcccggc	2100
tgatctgata	agtgtccgac	gagaagttct	gatggctttg	aagtatttca	catctggaga	2160
aatgaagtcg	aaattctttc	caatgctacc	tgcactcatc	gctgaggagg	ttgttctggg	2220
aacagatttc	actgcgattg	agcattttgcg	agttttcatg	tatcaaatgc	tagcagatct	2280
gttgcatcac	atgcgaaatt	ctatagacta	tgaaatgata	acacagtaag	tttgaataag	2340
actttctgat	gaaaaatgtt	gaaatttcag	cgtgattttc	gtattctgtc	gcactcttca	2400
cgatcctaac	aactcttctc	aagtcagat	tatgtctgct	cggctgctca	actcactggc	2460
cgaatctctg	tgcaaaatgg	attcacatga	taccgtaaga	cttattctat	caataatcgt	2520
atctcacttc	gaaataagtt	tcagactcgt	gatctgctca	ttgaaatcct	ggagtcgcac	2580
gtggccaagc	tcaaaactct	tgcagtctat	cacatgccta	ttctcttoca	acaatacggg	2640
accgaaatag	actacgaata	caaaagttat	gagagagacg	ccgagaaacc	tggaaatgaat	2700
atcccaaagg	acactatacg	aggagtaccg	aaacgaagaa	tccgtcggct	ctccattgat	2760
tcagttgaag	agctggaatt	cctggcatca	gaacctcca	cgtcgggaaga	tgcagatgag	2820
agtggtggag	atccgaacaa	gcttcctccg	ccaacaaaag	agggaaagaa	aacgtctccc	2880
gaagcgattt	taaccgccat	gtcaacgatg	acacctctc	cattggcaat	tgttgaagct	2940
cgaaatcttg	tgaagtatat	aatgcatacg	tgtaaattcg	tgacaggaca	attgagaatc	3000
gcccggccat	cacaggatat	gtatcattgt	tcgaaggagc	gagatttatt	cgaacgtctt	3060
ctacgatatg	gtgtaatgtg	tatggatgta	ttcgtgcttc	caacaactcg	aatcaacca	3120
caaatgcatt	cttcaatgcg	gacaaaagat	gagaaagatg	ctctggagtc	gttggcaaac	3180
gttttttcaa	caatcgacca	tgcatatctc	cgggaaatct	tcgaaaagta	tatggatttc	3240
ttgattgaaa	gaattttcaa	tcggaactat	ccattgcaat	tgatggtgaa	caccttcttg	3300
gttcgaaatg	aagtgccatt	cttcgcactc	acgatgcttt	cattcttgat	gtctcgaatg	3360
aaattgctgg	aagttagcaa	tgacaagacg	atgctatatg	tgaagctctt	caaaattatc	3420
ttctccgcca	tcggagccaa	tggtcttggt	cttcatggag	ataaaatgct	cacttcatac	3480
ctcccagaga	ttctcaaaca	gtcaactgtc	ttggcattaa	cagctcgtga	acctctcaac	3540
tatttctctt	tgcttcgtgc	attgttccgc	agtattgggt	gtggcgctca	ggatattttg	3600
tatggaaagt	tcctgcagtt	actgccaaat	cttcttcaat	tcttgaataa	attgacggtg	3660
agtttcattt	tttgatatat	cggtaataca	ctaaaaatcc	agaatcttca	gtcatgtcaa	3720
catcggattc	aaatgcgtga	gctcttcgtc	gagttgtgtt	tgactgtgcc	agttcgactc	3780
agttcccttc	tgccatacct	accgcttctg	atggatccac	tggtgtgtgc	gatgaatggg	3840
agtcggaaca	tagttacaca	aggattgaga	acattggaat	tatgtgtgga	taacttgcaa	3900
cctgaatatc	ttctcgaaaa	tatgcttctc	gtccgtggag	ctttgatgca	aggcctctgg	3960
cgtgttgtat	cgaaagctcc	agatacatca	tcgatgacag	cagcgttcag	gatcctcgga	4020
aagttcggag	gagccaatcg	aaaacttctg	aatcaaccgc	aaattcttca	agtagccact	4080
ttaggcgacg	taagtttatt	tagtttatct	tcttctctgt	tttaagttct	aacattgatc	4140
ctattaacag	actgttcagt	cgtacatcaa	tatggaattc	tcgcggatgg	gactcgatgg	4200
caatcacagc	attcacctgc	cactgtccga	gttgatgaga	gtcgttgccg	atcagatgag	4260
atatccagct	gatatgatcc	ttaatccaag	tcctgcaatg	atcccgtaaa	ctcatatgaa	4320
gaaatggtgt	atggaattgt	cgaaagccgt	cttggttagcc	ggacttggat	cttcaggaag	4380
cccaattact	ccaagtgcaa	atcttccgaa	gattatcaag	aaacttcttg	aagattttga	4440
tccaaacaat	cgtaccactg	aagtatacac	atgtccgagg	gaaagtgatc	gagagctttt	4500

tgtgaatgca	cttctcgcaa	tggtttgtaa	gttcttaagt	tcttttctct	ctaatacagat	4560
ctatatatta	aatttttcag	acggaatatg	gaataaagac	ggtttccggc	atgtctatag	4620
caaattcttt	atcaaagttc	tccgccagtt	tgcgttgatt	ggagtactcg	aatacattgg	4680
tggaaatgga	tggatgcgtc	atgcagaaga	ggaaggtgtt	ctaccattgt	gccttgactc	4740
gtctgttatg	gttgatgctc	tgattatttg	tctctctgaa	acatcgtcaa	gcttcatcat	4800
tgctgggtgc	atgtctcttc	gtcatatcaa	tgagactctc	tgccttacac	ttcccgatat	4860
tgatcaaatg	tcgaaagtgc	caatgtgcaa	atacttgatg	gagaaggtgt	tcaaattgtg	4920
tcacgggcct	gcttggtatg	caagatctgg	tggaaatcaat	gcaattggat	acatgatcga	4980
atcgtttcca	cgaaaatttg	ttatggactt	tgtgatagat	gttggtgatt	cgatcatgga	5040
agttattttg	ggaaactgtt	aagaaatatc	aagtggtact	gctgattctg	cataccgattg	5100
tctcaagaaa	atgatgcgag	tctatttcat	caaagaagaa	ggccaagaag	aggagaatct	5160
gacactcgcg	actatttttg	tgtctgcaat	ctctaagcat	tacttccaca	gtaatgaaag	5220
agtcagagaa	tttgcgattg	gtttaatgga	tcattgtatg	gttcactcaa	gacttgacc	5280
atcccttgat	aagttctact	atcgattcaa	ggagttcttt	gagccagaat	taatgcgggt	5340
gctcacaaca	gttccaacaa	tgctattggc	agacgcagga	ggaagtgttg	atggagtcca	5400
aaactatatg	ttcaactgtc	cggatgggtt	tgatttcgaa	aaagatatgg	acatgtacaa	5460
gcgatatttg	tcacatctgc	tggatattgc	acaaaccgat	acattttacct	taaaccaaaag	5520
gaatgccttc	aaaaaatgcg	agacatgccc	atcgcatttc	cttctcccat	tcccaatcac	5580
tacacatat	gattcaatgc	gagccagtgc	tctacagtgt	cttgatgatc	cgatgatcgc	5640
aatgaagaag	caatacatcg	acaagggaa	agagctgggt	gatgagcata	agatgataga	5700
gatcctcgca	cttcgcagct	ccaagatcac	agttgatcaa	gtctacgaga	gcgatgaatc	5760
ttggagacga	ttgatgacag	ttctattgag	agcagtcact	gacagagaaa	ctcctgaaat	5820
tgcggagaag	cttcactcct	cacttttgaa	ggctccacca	atatccacaa	tcactatcgc	5880
aacatttggg	gcttcttaca	taagaaatat	tagtggagca	ggagatgaca	gtgattcaga	5940
tcgtcatatt	tcgtacaacg	atataatgaa	gttcaagtgt	ctcgtggagc	tcaatccaaa	6000
gattctgggt	acaaaaatgg	cagtgaatct	cgcaaatcaa	atggttaaat	ataagatgag	6060
tgacaagatc	tctaggattt	tgctagttcc	cagtagcttc	actgaagagg	agctcgatga	6120
tttgcgaagc	gagaagatga	aaggaattcg	agagttggat	atgattgggt	atacgggtta	6180
aatgtctgtc	ggatgcccg	tgaccacatt	cacggagcaa	attattgtgg	atatcagtcg	6240
ttttgctgct	cattttgagt	atgcttattc	gcaagatgta	cttgtaaatt	ggattgatga	6300
tgtcacagta	atcctcaaca	aaagtcccaa	agatgtatgg	aagttcttct	tgtctcgaga	6360
atcaattcta	gatcctgcac	gcagatcctt	tattcgaaag	atcatagtct	atcaatcaag	6420
tggtccactg	cgacaggaat	tcattggatac	tccggaatat	tttgagaaac	tcattgatct	6480
tgacgatgag	gagaataagg	atgaagatga	gagaaaaatc	tgggatcgtg	atatgtttgc	6540
attttgcgatt	gtcgatcgta	tctcgaagag	ctgccctgag	tggcttattt	ctccgaattc	6600
cccaattcca	agaattaaga	agttgttctc	cgaaacggaa	ttcaatgagc	gatatgtggt	6660
tcgagcattg	actgaggtga	agaaatttca	agaagagatc	atagtgaac	ggatgacaga	6720
gcacaagtac	aaggttccga	agctgattct	gaataccttc	ctgagatatt	tgaggttaatt	6780
tcaagatagt	ttgtaaaaat	taattacaaa	gaaatatatac	aaaactgaac	cccaaaaaaa	6840
aatttttgaa	tttcggatca	aaaaaattta	atattttctc	gaaaaatcct	tcaaaataac	6900
aaaaaattcg	aattctcact	tctaaaatta	tttttgaatt	tttaaataat	ttttgaacat	6960
ttctctatga	aattcatgtt	ttgggcctat	ttcaggctat	aaaaattatt	tttctgattt	7020
taaataactt	gcaaatttca	ggctcaacat	ctatgactac	gatctattca	tcgttatcgc	7080
ctcgtgtttc	aatggcaatt	tcgtcaccga	tctctctttt	cttcgcgaat	atcttgaaac	7140
tgaagtcatc	ccgaaagtgc	cgttacaatg	gcgagagag	ctgtttcttc	gaattatgca	7200
gaagtttgat	acggatccac	aaactgctgg	aacaagtatg	cagcatgtga	aggeccttca	7260
atatttgggt	attcccacgt	tgcatggggc	gttcgagcga	tatgatacgg	atgaaattgt	7320
tggcaccgca	ccaatagatg	attcggattc	ttcgaatggat	gtagatccgg	caggcagctc	7380
ggataacctt	gtggctcggt	taacatcagt	cattgattct	catcgtaatt	atctgagcga	7440
tggaaatggc	attgttttct	atcaactttg	cacattgttc	gtacaaaacg	cctccgaaca	7500
tattcacaat	aataactgca	agaaacaagg	tggacgccta	cggatcctga	tgtcttctgc	7560
ctggccgtgc	ctgaccatgt	acaatcatca	agatccaaca	atgcggtaca	ctggattctt	7620
cttcttggcc	aattattatg	agcgtttcac	aattaatcgg	aaaatcgtgc	ttcaagtgtt	7680
ccatcaactt	atgactactt	atcagcagga	cactagagat	caaatccgga	aagccattga	7740
tatattaact	ccagctttga	ggacacgaat	ggaagatgga	cacttgcaaa	tattgagtca	7800
tgtgaagaaa	attcttatcg	aagaatgcc	taatttgcaa	catgttcagc	atgttttgta	7860
agtttattat	ctaaaaatgat	tttttttaatt	gttaaaaaatt	taatttttaa	atgcgttcgt	7920
gctcctttta	taattcctga	attttccagc	caaatgggtg	ttcgcaatta	tcgtgtctac	7980

tatcatgttc	gattggagct	tctcacgcct	cttctgaacg	gagttcaacg	agcacttgtg	8040
atgccaaata	gtgttctgga	aaagtaagtt	tccagcccgt	tgttcgtaaa	ctcaccctt	8100
gtaaatat	agctggcaaa	ctcgacgtca	tgcggtggag	atctgcgaga	tggtcatcaa	8160
gtgggaattg	ttcagaacgc	tgaaaacaga	tcatattatc	agtgcgaag	aagctctcga	8220
agttgacaag	caattggata	agctgcgaac	agcttcatcc	acagatcggt	tcgatttcga	8280
ggaggctcat	aacaagagag	acatgcctga	tgctcaacgc	acgattatca	aagagcacgc	8340
cgatgtgatt	gtcaatatgc	ttgtccgatt	ctgtatgacg	ttccatcaga	attcggttcc	8400
ttcgtccact	tctcaaagtg	ggaaccatgg	tgtcgagttg	acaaaaaat	gtcagctgct	8460
tctacgtgca	gccctacgac	caagcatgtg	gggagaattt	gtcagcttcc	gattaacaat	8520
gatcgaaaag	tttttgtcaa	ttccgaatga	taatgctcta	cgcaatgata	taagttctac	8580
ggcctacgct	aatactatcc	aaaatgcaca	acacactctg	gatatgctgt	gtaatattat	8640
tcctgttatg	ccaaaaacta	gcttgatgac	tatgatgaga	caactccaac	ggccactcat	8700
acaatgtctc	aataacggag	ctcaggtatg	tgaagaacga	tgaatagggg	gttataaatc	8760
actaatttct	cttagaactt	taagatgact	cgtcttgtca	ctcaaattgt	cagtcgggta	8820
ctcgaaaaga	caaagtgttc	ggttaacggg	cttgatgagc	tgagagcaatt	gaatcaatac	8880
atttcccgat	tcctacatga	acatttttga	tctcttttga	agtaagtttt	atttttgaat	8940
ttccatcttt	caacccttcg	ccagttgcag	aaactttgag	ggaccagtgt	tggagttctc	9000
cggagcattt	tctcttttgc	gaacaatttg	tggacacgag	ccagcatact	tggcattt	9060
gatgccttca	tttgtaaaag	tgatggagag	agctgcaaaa	gagcacttgg	cgtatgttgc	9120
gaactcgcaa	gatggaaata	tggatgaagag	taagttctat	aaaaagattc	agattttcta	9180
atccccttag	atttctttcc	agatgttgct	gaattgttgt	gtgcatgcat	ggagctggta	9240
cgtcccagag	tcgatcatat	cagtatggag	attaagagat	caattgttgg	tggattatc	9300
gcggagctga	ttatcaaata	gaatcacgat	aagatcatcc	agacgtcagt	gaagcttctc	9360
ggagcaatga	ttagcacgca	ggatatggaa	tttacaattc	tcaactgttct	tccgctactt	9420
gttcgtatcc	aatcaattat	tgtgaccaag	ttcaagaatt	gcaaggatct	gatagcagac	9480
tatcttgttg	tggttattac	cgttttttga	aacagcgaat	atcggaactc	ggaagctgga	9540
tctcgtctct	gggaaggatt	cttctgggga	ctcaagagta	gcgacctca	aaccgggag	9600
aaattctcga	tagtttggga	gaagacttgg	ccacacatgg	caacagtaga	tattgtctcat	9660
cgaatgaaat	atatcatgca	aaatcaagat	tgggtccagt	tcaaacacgc	gttttgggtg	9720
aaattcgcac	tttggggaat	gctacgaacg	attgccaaac	ggccaactga	tccgaataat	9780
aagagaaaga	aagtgatact	gttgaactgt	gcaactccat	ggagaacaat	tgaatatgca	9840
gcgaaattga	aggatcagcc	aatggaagtg	gaaactgaaa	tgaacgaga	agagccagaa	9900
ccgatggaag	ttgacgaaaa	agactcgcaa	gatgattcta	aggatgccgg	agagcccaag	9960
gagaaggaaa	agctcacatt	ggaattattg	cttgctggac	aacaagaact	tttggatgaa	10020
gcttccaatt	atgatttttgc	ggatgctcta	gatacagtat	cccagattac	atttgcactt	10080
aatggtaaat	tgttcaaagt	ttatgaatat	ttttcttaaa	aatcacaatt	ttcagagaaa	10140
caagtgcaca	gcaagatgtg	ggtagtgttg	ttcaaatacat	tctggagtcc	cttatcaca	10200
tccgaaatcg	aagattttcac	ggcgctagtc	gttccgttta	tgagcagtgg	agtgcataat	10260
aattatcaga	cgggtgtaca	ggatagtgtg	cttgctgttt	ggcttgaagc	tgttggtgac	10320
gctgttcatt	tgccgtccag	attgattgag	gtacgttctg	aaaatgaatg	ctggaaaaaa	10380
ttcgattttt	ctgttttaaaa	aaagttaaaa	tttccgattt	tttgaatagc	aaaaaaaaaa	10440
gaaaacattt	attttgaaaa	aagagtcctc	accggaattt	tttaataaat	aaatttaaaa	10500
aaagaaaaaa	aactaaaaac	ttcaattttt	gaaaaatcaa	aaaaaaatta	cagaaacaga	10560
cgaggtaaaa	aatttttaaaa	aagttctgta	aaaaaaatgg	agaatcacag	ttttcgttgt	10620
cttttctgaa	aaaaatttga	aaaatttaaaa	attaacgatt	ttttggtttt	taatttaaaa	10680
aaatatacga	aaaaagactg	aagaactttt	tttgtcaaaa	aaacttgatt	ttgatgaggg	10740
aaaaagttca	aaaacttgga	gaaatcatcg	gaaatttttag	aagattcaat	aaaaatttcc	10800
aaaaaaaaaa	attgaacatt	tatgattttt	gggtattttt	aaaaattgaa	aaattacgct	10860
taatttttag	attaaaaaaa	tcaaaaaaaa	accaacactc	cttttgaaac	ttgacacttt	10920
tgaacgttt	tttttttttg	caataataaa	tttctcattt	cagtttatct	catcaaaaac	10980
cgaatgctgg	cataccggaa	tcaggcttct	cgagaatcat	atatggacaa	ttccaaagca	11040
actcaacaac	acgttactcc	gagaaatgaa	agtggcacca	ggtctcgctg	gagatattga	11100
gacactcgaa	tctcttggaa	cactctacaa	tgagatatca	gagtttgatc	agttcgctgc	11160
aatctgggaa	cgccgtgctg	tatttctctga	tacgatgaga	gcaatgtcag	ctatgcaatt	11220
gggagatag	gaattagctc	aatcttatct	ggaaaaatca	atgagcagta	cgtatgaaac	11280
tcttgctccg	acaatcaatc	gtaagtttgg	atcaatcggt	tgtacttctc	acacaaaaata	11340
gtattccttt	cagcaaaaca	cacttcaaat	tcggagaagc	atgtttctcc	gattattgac	11400
aaagaatacg	atcattggat	ggagatgtac	atcacaaatt	gctcggagct	tcttcagtgg	11460

caaaatgtgg	ccgacgtatg	caatggcaaa	gacatgcaac	atgttcgtgg	cctgatcaac	11520
gcagcatctc	acattccgga	ctggaatgtg	gtcgaaggagt	gtaaaagtca	gatagctgga	11580
tgtattccac	caagtttcca	tttagattac	actcttttca	atttgatgag	tactgttatg	11640
gttagtttaa	gtcaaaaagt	gatatataat	tattgtttaa	tttttcagcg	aatgaatgaa	11700
aactcaagcc	cgacacatat	gaaggaacga	tgcaaaattg	caattcaaga	gtgcacagaa	11760
gctcatatta	gtcgttggag	agcacttccg	tcagttgttt	catatggtca	tgtcaagatt	11820
cttcaggcaa	tgaacttggg	tcgagaaatt	gaagagtcta	cagatattcg	cattgctctg	11880
ctcgaggccc	catcaaacia	agtggatcag	gcgttgatgg	gcgatatgaa	gtcgttgatg	11940
aaagtattcc	gaaatagaac	accaaccact	tcggatgata	tgggattcgt	ttcgacttgg	12000
tatgattgga	ggaatcagat	tcattggaat	atgcttcaaa	gattcgaata	ttgggataaa	12060
gtaggactca	acgtcgctgc	aactggaaaac	cagtcaattg	ttccgattca	ttcaatggct	12120
caagcacagt	tggccgtagc	caaacatgcc	aagaatcttg	gattccataa	tttaacgaaa	12180
gatctactca	acaaatttagc	tggattgaca	gccataccga	tgatggatgc	tcaagataaa	12240
gtttgcactt	acggcaagac	acttcgcgat	atggcaaaac	gtgcggctga	cgaagagtg	12300
aaaaatgagc	tattgtgtga	agcgcttgaa	gttttggaag	atgtgcgaat	tgatgatcta	12360
cagaaggatc	aggttgctgc	attgctttat	catcgtgcta	atattcattc	agttcttgat	12420
cagtaagttt	tcaatgccga	aaaaaaatta	aagttttaca	aaaataaatt	tcagagctga	12480
aatgctgac	tacaccttct	ccgcagcctc	tcaacttgct	gacttgcaaa	atagtgtgac	12540
aaccactgga	atcaagctca	tgaaaaattg	gggccaccat	ctttacaaga	gattcttctc	12600
tacgacagtt	tgcaaggaaa	ccggaacaaa	cttcggacgg	caggctctcg	cttggtactt	12660
cattgcggct	cgtgtggata	acgatatcaa	ggcgagaaa	ccgattgccca	agattttctg	12720
gctctcgaag	cacttgaatg	cgtgtggatc	acatgaagtg	atgaatcggg	ttattaagaa	12780
gcaacttcat	tcacttaatc	tcttcaattg	gctttactgg	cttcacacaa	tggttactga	12840
tgttcgatat	aaaccaaatt	cgaactttgt	tctgattctc	tgcaaggtaa	gttttgaaat	12900
atttaaatat	tttcagaatt	ttaaatgaaa	ttcatttgca	gatggctgct	gctcatccac	12960
ttcaagtatt	ttaccacatt	cgaggaggcag	ttagcgttga	cgatattgac	tcggttctcg	13020
aagaagatta	cactgatgag	caaatgtcga	tggatgtttc	ggatgaggat	tgttttgag	13080
acgatccacc	atttgataga	attctgaaaa	tatgtctgaa	atatcgtcca	actgatattc	13140
gagtcttcca	tcgtgtcctc	aaagaacttg	acgatagtaa	tgagacatgg	gttgaacgtc	13200
acttgcgtca	tcgatctgc	ctcaaggatc	agatgttcaa	agatttctcg	gaacaaatgg	13260
acgcgacggt	caatgagatg	caatattcgg	aggatgtgac	tatgatgacg	ttgagatgga	13320
ggaaacagct	ggaagaagac	ttggtgtatt	tccaacagaa	ttataatctt	gatttcctgg	13380
agattcgtaa	caagcgaaag	atgatcgtga	cgaagggatg	tatgggagtc	gagaaaagtc	13440
agataatggt	cgaaaaagag	ctgagtcga	tgttcacaga	gccggccggc	atgcaagatg	13500
aatttgattt	tgtcacaaat	atgactaata	tgatggtctc	acagttggat	attcatgcag	13560
tcgatgctcc	acgccctcag	ggatatattc	gtattgttct	cgactggatt	cgagcgattc	13620
gtcgtcggtt	cgatcgactt	ccacgaagaa	tcctcttgga	atcgtcaagc	ccatctctcg	13680
ccagattcag	ccatcgatca	ggatgcatcg	aaatgccata	cgatttgctc	aacgttttgc	13740
gcgccaagaa	tcataactctg	atggcttcca	atcaaacggg	gcaatacata	tccatgctct	13800
ctcgatttga	gccaaacttt	gagattgtga	tcaaagggtg	tcaagtgata	agaaagatct	13860
atattcagag	acaaaccgga	aagagtgcgg	cgttttatct	gaagaaatct	gtgcaggatg	13920
agccaactaa	ccgagttcca	caaattgttc	aacatcttga	tcacgttcta	caaaccgata	13980
gagagtcggc	gagaagacat	cttcatgctc	caacagtgc	gcagatgaga	gtcggacaga	14040
agacgacact	ctacgaagtt	gcatccgttc	aacctatg	aatgccaccg	gattgtacca	14100
gaaactatcc	agcatcacaa	atcgacattg	ttcatccata	tgatgtgctg	actgccactt	14160
tcaatggaag	ttattatccg	gatgatattg	tattgcactt	ctttgagaga	ttcgcccaaa	14220
gttcttcatc	catcggaaca	cctcttccaa	ctccgacgaa	ccaagatgga	acagttgctc	14280
cgccacgact	aacggaagct	caccacatca	agaatattat	ttatgagtac	gtttgagaag	14340
ctagtgtcta	aaataataat	taatgtaaaa	aaattttcag	agactttgcc	cgagatatga	14400
tccattccg	acttctctac	gactacctca	ctgcacgata	tcctgatccg	gttatgtact	14460
atgcaatgaa	gaagcaattg	ctgcacagtc	tcgccgtcct	atccacaatc	gaatatcatt	14520
gcaatctgac	accaatggga	cctgatcaaa	tgatgatgac	aatgaatact	ggagtcctta	14580
gcaatccttc	atatagattc	gaaatccgag	gaggacgatc	acttcatgat	attcaacact	14640
ttggacatga	agttccattc	cgattgactc	caaactctatc	gattttgggt	ggtgttgac	14700
aggatggtga	cttggttatgg	agtatggctg	ctgcgtcaaa	atgtttgatg	aagaaggaa	14760
ctgaagttat	catgagaccg	ttagtatggg	atgaattcgc	caacaatata	gattgcgaca	14820
aatcggtaat	tttactttaat	tatgctaata	gggaattgaa	ctaattgttt	ccaagcggtt	14880
gcagggtattc	gcgtgtcatg	catcgaattc	ttacatcaat	ggtgtcgcga	gcaagcttcg	14940

```

aaacacgaat agcgccgacg ccaaactcag aaaggacgat tgtgtgtcgc tgatcagtcg 15000
agccaaggat tcggataatc tggcccgaat gcccccacc taccacgcgt gggtctagat 15060
ctcataatta ccgttctcta ttttgatccc gcctcccact ctacagatc tctatacatt 15120
tgtcaaatgt ttccaaatct tttatctgcc catacattcg tttttattgt tttgtttctt 15180
ttctttcttt atttcttttc taaactttta gatttatgta aatatttaac tgcgctggta 15240
tttatgaaaa attcagataa agttttcaag tttaaaaaat cgaaaattcg aagtcggaag 15300
ttctcttaca ggtgtagtaa gtaggcacaa tggcaatagg tacatggaag gcttgcgga 15360
ggcacatggg taggcataag atcgaaaaat aagctgatat ataaatatag ataggtattg 15420
gttaggcaca aattaggcac gtaggtgtga gctggcaaat aggtaggcat gacgttcggc 15480
aaatcgcaa attgccgatt tggcgaaaat tttcaaacc gcgatttgc cggaaatgtt 15540
tagagaaatt ttttataaga cagaaaaact tacaactgtg tctttttgaa attcttcgg 15600
ttttctttat acagtgcgtg caacttctat agcgcccccc ccccccccc cccccctat 15660
tttttcgcat ttcacgccat tctgattttt atttttctga ttttttttt tttgcaactga 15720
aacttgcat tgaggatgct tggagagaaa tatcagccag caaaataaag aatctggtca 15780
actcaatgct gaatagattt tttgaggtta tcgttaagaa gggaggtccc acgacgtatt 15840
gatccttcat cgagttaaca aattatgatg ttttaattga tttcattcca cttctggaca 15900
cagaaggacg aatagtgaac tctggtacaa gtttatcacc acctacaact tcgtcgattt 15960
gtggaaaatc tttcagacat gtctccatga gtgtctcaga acatcttggt caggtttgga 16020
gtcgatccca ccgctgggag ccgagaatgg gcctctaaca c 16061

```

<210> 13

<211> 12195

<212> DNA

<213> *Caenorhabditis elegans*

<400> 13

```

atggatccgg ctatggcttc tccaggctat cggctctgtgc agtccgatcg gagtaatcac 60
ctaacagagc tggaaacgag aattcaaaat cttgccgata attcacaag agatgatgct 120
aaattgaaaa tggtacaaga gatttggagc acaatcgaaa atcatttcac actaagttcg 180
cacgagaaag tcgtggagag gctcattctc tcgttccctac aagttttctg caacacaagt 240
ccacagttca ttgctgaaaa caatacacaa cagcttcgaa agttaatgct tgaaatcatt 300
cttcgacttt cgaacgtaga agccatgaaa catcatagca aagaaattat caagcagatg 360
atgaggctaa tcaccgtgga aaatgaggag aatgccaat tggctatcaa aattgtcacc 420
gatcaaggga gaagtaccgg caaaatgcaa tattgctggag aggtttcaca gataatggct 480
tccttcaaaa caatggtcat tgatctgacg gcgagtggtc gagctgggtga tatgttcaac 540
ataaaagagc ataaagctcc accgtcaact agctccgacg agcaagtcac cactgaatat 600
ttgaagactt gctactatca acaaaccggt cttctcaacg gaacggaagg aaaaccgcca 660
ttaaaatata atatgattcc atcagctcat cagtcacga aggtgctcct ggaggttccg 720
tatctcgtag ttttcttcta tcaacatttc aaaacagcga tccaaaccga agcgcttgat 780
ttcatgaggg ttggtcttga ttttctaaat gtcagagttc cagacgagga taaactcaaa 840
acaaatcaaa taataaccga tgattttgtc agtgcacagt cccgattcct gtcattcgtc 900
aacattatgg ctaagattcc agcgtttatg gatcttatca tgcaaaatgg accgcttcta 960
gtgtcgggaa caatgcagat gctcgagcgg tgcccggctg atctgataag tgtccgacga 1020
gaagttctga tggctttgaa gtatttcaca tctggagaaa tgaagtcgaa attctttcca 1080
atgctacctc gactcatcgc tgaggaggtt gttctgggaa caggattcac tgcgattgag 1140
catttgcgag ttttcatgta tcaaagtcta gcagatctgt tgcatacat gcgaaattct 1200
atagactatg aaatgatcac acacgtgatt ttogtattct gtgcactct tcacgatcct 1260
aacaactctt ctcaagtcca gattatgtct gctcggtgc tcaactcact ggcgaatct 1320
ctgtgcaaaa tggattcaca tgataccttt cagactcggt atctgctcat tgaaatcctg 1380
gagtcgcacg tggccaagct caaaactctt gcagctctat acatgcctat tctcttccaa 1440
caatacggaa ccgaaataga ctacgaatac aaaagttatg agagagacgc cgagaaacct 1500
ggaatgaata tcccaaagga cactatacga ggagtaccga aacgaagaat ccgtcggctc 1560
tccattgatt cagttgaaga gctggaattc ctggcatcag aaccatccac gtcggaagat 1620
gcagatgaga gtggtggaga tccgaacaag cttcctccgc caacaaaaga gggaaagaaa 1680
acgtctcccg aagcgatttt aaccgccatg tcaacgatga cacctcctcc attggcaatt 1740
gttgaagctc gaaatcttgt gaagtatata atgcatacgt gtaaatctgt gacaggacaa 1800

```

ttgagaatcg	cccgccatc	acaggatatg	tatcattggt	cgaaggagcg	agattttatc	1860
gaacgtcttc	tacgatattg	tgtaatgtgt	atggatgtat	tcgtgcttcc	aacaactcga	1920
aatcaaccac	aaatgcattc	ttcaatgcgg	acaaaagatg	agaaagatgc	tctggagtcg	1980
ttggcaaacg	tttttacaac	aatcgaccat	gcgatattcc	gggaaatctt	cgaaaagtat	2040
atggatttct	tgattgaaag	aattttacaat	cggaactatc	cattgcaatt	gatgggtgaac	2100
accttcttgg	ttcgaaatga	agtgccattc	ttcgcatcta	cgatgctttc	attcttgatg	2160
tctcgaatga	aattgctgga	agtttagcaat	gacaagacga	tgctatatgt	gaagctcttc	2220
aaaatttatct	tctccgccat	cggagccaat	ggctctgggc	ttcatggaga	taaaatgctc	2280
acttcatacc	tcccagagat	tctcaaacag	tcaactgtct	tgccattaac	agtcctgtaa	2340
cctctcaact	atttctttt	gcttcgtgca	ttgttccgca	gtattgggtg	tgccgctcag	2400
gatattttgt	atggaaagtt	cctgcagtta	ctgccaaatc	ttcttcaatt	cttgaataaa	2460
ttgacgaatc	ttcagtcatt	tcaacatcgg	attcaaagtc	gtgagctctt	cgctgagttg	2520
tgtttgactg	tgccagttcg	actcagttcc	cttctgccat	acctaccgct	tctgatggat	2580
ccactggtgt	gtgcgatgaa	tgggagtcgg	aacatagtta	cacaaggatt	gagaacattg	2640
gaattatgtg	tggataactt	gcaacctgaa	tatcttctcg	aaaatatgct	tcctgtccgt	2700
ggagctttga	tgcaaggcct	ctggcggtgt	gtatcgaaaag	ctccagatac	atcatcgatg	2760
acagcagcgt	tcaggatcct	cggaaaagttc	ggaggagcca	atcgaaaact	tctgaatcaa	2820
ccgcaaattc	ttcaagtagc	cacttttaggc	gacactgttc	agtcgtacat	caatatggaa	2880
ttctcgcgga	tgggactcga	tggcaatcac	agcattcacc	tgccactgtc	cagtttgatg	2940
agatcgtttg	ccgatcagat	gagatatcca	gctgatatga	tccttaatcc	aagtcctgca	3000
atgatcccg	caactcatat	gaagaaatgg	tgtatggaat	tgtcgaaagc	cgctctgtta	3060
gccggacttg	gatcttcagg	aagcccaatt	actccaagtg	caaactcttc	gaagattatc	3120
aagaaacttc	ttgaagattt	tgatccaaac	aatcgtacca	ctgaagtata	cacatgtccg	3180
agggaaagtg	atcgagagct	ttttgtgaat	gcacttctcg	caatggctta	cggaatatgg	3240
aataaagacg	gtttccggca	tgtctatagc	aaattcttta	tcaaagttct	ccgccagttt	3300
gcgttgattg	gagtactcga	atacattggt	ggaaatggat	ggatgctgca	tgcagaagag	3360
gaaggtgttc	taccattgtg	ccttgactcg	tctgttatgg	ttgatgctct	gattatttgt	3420
ctctctgaaa	catcgtcaag	cttcatcatt	gctggtgtca	tgtctcttcg	tcatatcaat	3480
gagactctct	cgcttacact	tcccgatatt	gatcaaatgt	cgaaagttcc	aatgtgc aaa	3540
tacttgatgg	agaaggtgtt	caaattgtgt	cacgggcctg	cttggtatgc	aagatctggg	3600
ggaatcaatg	caattggata	catgatcgaa	tcgtttccac	gaaaatttgt	tatggacttt	3660
gtgatagatg	ttgttgattc	gatcatggaa	gttattttgg	gaactgttga	agaaatatca	3720
agtggatctg	ctgattctgc	atacgattgt	ctcaagaaaa	tgatgagagt	ctatttcatc	3780
aaagaagaag	gccaagaaga	ggagaatctg	acactcgcga	ctatttttgt	gtctgcaatc	3840
tctaagcatt	acttccacag	taatgaaaga	gtcagagaat	ttgcgattgg	tttaatggat	3900
cattgtatgg	ttcactcaag	acttgacca	tcccttgata	agttctacta	tcgattcaag	3960
gagttctttg	agccagaatt	aatgcgggtg	ctcacaacag	ttccaacaat	gtcattggga	4020
gacgcaggag	gaagtttgga	tggagttcaa	aactatatgt	tcaactgtcc	ggatgggttt	4080
gatttcgaaa	aagatatgga	catgtacaag	cgatatttgt	cacatctgct	ggatattgca	4140
caaaccgata	catttacctt	aaaccaaagg	aatgccttca	aaaaatgcga	gacatgcccc	4200
tcgcatttcc	ttctccatt	cccaatcact	acacatattg	attcaatgcg	agccagtgtc	4260
ctacagtgtc	ttgtgatcgc	gtatgatcga	atgaagaagc	aatacatcga	caagggataa	4320
gagctgggtg	atgagcataa	gatgatagag	atcctcgcac	ttcgcagctc	caagatcaca	4380
gttgatcaag	tctacgagag	cgatgaatct	tggagacgat	tgatgacagt	tctattgaga	4440
gcagtcaactg	acagagaaac	tcttgaaatt	gcggagaagc	ttcatccttc	acttttgaag	4500
gtctcaccaa	tatccacaat	catcatcgca	acatttggtg	cttcttacat	aagaaatatt	4560
agtggagcag	gagatgacag	tgattcagat	cgtcatattt	cgtacaacga	tataatgaag	4620
ttcaagtgtc	tcgtggagct	caatccaaag	attctggtca	caaaaatggc	agtgaatctc	4680
gcaaatcaaa	tggttaaata	taagatgagt	gacaagatct	ctaggatttt	gtcagttccc	4740
agtagcttca	ctgaagagga	gctcgatgat	ttcgaagcgg	agaagatgaa	aggaattcga	4800
gagttggata	tgattggtca	tacggttaaa	atgcttgctg	gatgccaggt	gaccacattc	4860
acggagcaaa	ttattgtgga	tatcagtcgt	tttgctgctc	attttgagta	tgcttattcg	4920
caagatgtac	ttgtaaattg	gattgatgat	gtcacagtaa	tcctcaacaa	aagtcccaaa	4980
gatgtatgga	agttcttctt	gtctcgagaa	tcaattctag	atcctgcacg	cagatccttt	5040
attcgaagaa	tcatagtcta	tcaatcaagt	ggtccactgc	gacaggaatt	catggatact	5100
ccggaatatt	ttgagaaact	cattgatctt	gacgatgagg	agaataagga	tgaagatgag	5160
agaaaaatct	gggatcgtga	tatgtttgca	ttttcgattg	tcgatcgtat	ctcgaagagc	5220
tgccttgagt	ggcttatttc	tccgaattcc	ccaattccaa	gaattaagaa	gttgttctcc	5280

gaaacggaat	tcaatgagcg	atatgtggtt	cgagcattga	ctgaggtgaa	gaaatttcaa	5340
gaagagatca	tagtgaaacg	gatgacagag	cacaagtaca	aggttccgaa	gctgattctg	5400
aataccttcc	tgagatat	gaggctcaac	atctatgact	acgatctatt	catcgttatc	5460
gcctcgtgtt	tcaatggcaa	tttcgtcacc	gatctctctt	ttcttcgcga	atatcttgaa	5520
actgaagtca	tcccgaaagt	gccgttacaa	tggcggagag	agctgtttct	tcgaattatg	5580
cagaagtttg	atacggatcc	acaaactgct	ggaacaagta	tgcagcatgt	gaaggccctt	5640
caatatttgg	ttattcccac	gttgcattgg	gcgttcgagc	gatatgatac	ggatgaaatt	5700
gttggcaccg	caccaataga	tgattcggat	tcttcgatgg	atgtagatcc	ggcaggcagc	5760
tcggataaac	ttgtggctcg	tttaacatca	gtcattgatt	ctcatcgtaa	ttatctgagc	5820
gatggaatgg	tcattgtttt	ctatcaactt	tgcacattgt	tcgtacaaaa	cgcctccgaa	5880
catattcaca	ataataactg	caagaaacaa	ggtggacgcc	tacggatcct	gatgctcttc	5940
gcctggcgtg	gcctgaccat	gtacaatcat	caagatccaa	caatgcggtg	cactggattc	6000
ttcttcttgg	ccaatattat	agagcgtttc	acaattaatc	ggaaaatcgt	gcttcaagtg	6060
ttccatcaac	ttatgactac	ttatcagcag	gacactagag	atcaaatacg	gaaagccatt	6120
gatatatata	ctccagcttt	gaggacacga	atggaagatg	gacacttgca	aatattgagt	6180
catgtgaaga	aaattcttat	cgaagaatgc	cataatttgc	aacatgttca	gcattgtttc	6240
caaattggtg	ttcgcaatta	tcgtgtctac	tatcatgttc	gattggagct	tctcacgcct	6300
cttctgaacg	gagttcaacg	agcacttggt	atgccaaata	gtgttctgga	aaaatttagc	6360
tggcaaac	gagtcacgc	ggtggagatc	tgcgagatgg	tcatacaagt	ggaattgttc	6420
agaacgctga	aaacagatca	tattatcagt	gacgaagaag	ctctcgaaat	tgacaagcaa	6480
ttggataagc	tgcgaacagc	ttcatccaca	gatcgtttcg	atttcgagga	ggctcataac	6540
aagagagaca	tgcctgatgc	tcaacgcacg	attatcaaag	agcacgccga	tgtgattgtc	6600
aatatgcttg	tccgattctg	tatgacgttc	catcagaatt	cgggttcttc	gtccacttct	6660
caaagtggga	accatgggtg	cgagttgacc	aaaaaatgtc	agctgcttct	acgtgcagcc	6720
ctacgacca	gcatgtgggg	agaatttgct	agcttccgat	taacaatgat	cgaaaagttt	6780
ttgtcaattc	cgaatgataa	tgcctctacg	aatgatataa	gttctacggc	ctacgcta	6840
actatccaaa	atgcacaaca	cactctggat	atgctgtgta	atattattcc	tgttatgcca	6900
aaaactagct	tgatgactat	gatgagacaa	ctccaacggc	cactcataca	atgtctcaat	6960
aacggagctc	agaactttaa	gatgactcgt	cttgtcactc	aaattgtcag	tcggttactc	7020
gaaaagacaa	atgtttcggg	taacgggctt	gatgagctgg	agcaattgaa	tcaatacatt	7080
tcccgattcc	tacatgaaca	ttttggatct	cttttgaaat	gcagaaaact	gagtggacca	7140
gtgttgggag	ttctcggagc	attttctctt	ttgcgaacaa	tttgtggaca	cgagccagca	7200
tacttggatc	atttgatgcc	ttcatttgta	aaagtgatgg	agagagctgc	aaaagagcac	7260
ttggcgtatg	ttgcgaactc	gcaagatgga	aatatgggtg	agaatttctt	tccagatgtt	7320
gctgaattgt	tgtgtgcatg	catggagctg	gtacgtccca	gagtcgatca	tatcagtatg	7380
gagattaa	gatcaattgt	tgggtggtat	atcgcggagc	tgattatcaa	atcgaatcac	7440
gataagaatc	tccagacgtc	agtgaagctt	ctcggagcaa	tgattagcac	gcaggatatg	7500
gaatttacaa	ttctcactgt	tcttccgcta	cttgctcgta	tccaatcaat	tattgtgacc	7560
aagttcaaga	attgcaagga	tctgatagca	gactatcttg	ttgtgggtat	taccgttttt	7620
gagaacagcg	aatatcggaa	ctcgggaagct	ggatctcgtc	tctgggaagg	attcttctgg	7680
ggactcaaga	gtagcgatcc	tcaaaccggg	gagaaattct	cgatagtgtg	ggagaagact	7740
tggccacaca	tggcaacagt	agatattgct	catcgaatga	aatatatcat	gcaaaatcaa	7800
gattgggtcca	agttcaaaca	cgcgttttgg	ttgaaattcg	cactttgggg	aatgctacga	7860
acgattgcca	aacggccaac	tgatccgaat	aataagagaa	agaaagtgat	actgttgaac	7920
tgtgcaactc	catggagaac	aattgaatat	gcagcgaaat	tgaaggatca	gccaatggaa	7980
gtggaaaactg	aaatgaaacg	agaagagcca	gaaccgatgg	aagttgacga	aaaagactcg	8040
caagatgatt	ctaaggatgc	cggagagccc	aaggagaagg	aaaagctcac	atttgaatta	8100
ttgcttgctg	gacaacaaga	acttttggat	gaagcttcca	attatgattt	tgcggatgct	8160
ctagatacag	tatcccagat	tacatttgca	cttaatgaga	atcaagtgc	aagcaagatg	8220
tgggtagtgt	tgttcaaate	attctggagt	tccttatcac	aatccgaaat	cgaagatttc	8280
acggcgctag	tcgttccggt	tatgagcagt	ggagtgcata	ataattatca	gacgggtgta	8340
caggatagtg	tgcttgctgt	ttggcttgaa	gctgttggtg	acgctgttca	tttgccgtcc	8400
agattgattg	agtttatctc	atcaaaacac	gaatgctggc	ataccggaat	caggcttctc	8460
gagaatcata	tatggacaat	tccaaagcaa	ctcaacaaca	cgttactccg	agaaatgaaa	8520
gtggcaccag	gtctcgctgg	agatattgag	acactcgaat	ctcttggaa	actctacaat	8580
gagatgacag	agtttgatca	gttcgctgca	atctgggaac	gccgtgctgt	atttctctgat	8640
acgatgagag	caatgtcagc	tatgcaattg	ggagatatgg	aattagctca	atcttatctg	8700
gaaaaatcaa	tgagcagtac	gtatgaaact	cttgctccga	caatcaatcc	aaacaacact	8760

tcaaattcgg	agaagcatgt	ttctccgatt	attgacaaag	aatacgatca	ttggatggag	8820
atgtacatca	caaattgctc	ggagcttctt	cagtggcaaa	atgtggccga	cgtatgcaat	8880
ggcaaagaca	tgcaacatgt	tcgtggcctg	atcaacgcag	catctcacat	tccggactgg	8940
aatgtggtcg	aggagtgtaa	aagtcagata	gctggatgta	ttccaccaag	tttccattta	9000
gattacactc	ttttcaattt	gatgagtact	gttatgcgaa	tgaatgaaaa	ctcaagcccg	9060
acacatatga	aggaacgatg	caaaattgca	attcaagagt	gcacagaagc	tcatattagt	9120
cgttggagag	cacttccgtc	agttgtttca	tatggctcatg	tcaagattct	tcaggcaatg	9180
aacttggttc	gagaaattga	agagtctaca	gatattcgca	ttgctctgct	cgaggcccca	9240
tcaaacaaag	tggatcaggc	gttgatgggc	gatatgaagt	cgttgatgaa	agtattccga	9300
aatagaacac	caaccacttc	ggatgatatg	ggattcggtt	cgacttggtg	tgattggagg	9360
aatcagattc	atggaatgat	gcttcaaaga	ttcgaatatt	gggataaagt	aggactcaac	9420
gtcgctgcaa	ctggaaacca	gtcaattggt	ccgattcatt	caatggctca	agcacagtgt	9480
gccgtagcca	aacatgccaa	gaatcttgga	ttccataaatt	taacgaaaga	tctactcaac	9540
aaattagctg	gattgacagc	cataccgatg	atggatgctc	aagataaagt	ttgcacttac	9600
ggcaagacac	ttcgcgatat	ggcaaacagt	gcggtgacg	aaagagtga	aaatgagcta	9660
ttgtgtgaag	cgcttgaagt	tttgggaagt	gtgccaattg	atgatctaca	gaaggatcag	9720
gttgctgcat	tgctttatca	tcgtgcta	attcattcag	ttcttgatca	agctgaaaa	9780
gctgactaca	ccttctccgc	agcctctcaa	cttgctgact	tgcaaaatag	tgtgacaacc	9840
actggaatca	agctcatgaa	aaattggggc	caccatcttt	acaagagatt	cttctctacg	9900
acagtttgca	aggaaccggg	aaacaacttc	ggacggcagg	ctctcgcttg	ttacttccatt	9960
gcgggtcgtg	tggataacga	tatcaaggcg	agaaaaccga	ttgccaagat	tttgtggctc	10020
tcgaagcact	tgaatgcgtg	tggatcacat	gaagtgatga	atcggttat	taagaagcaa	10080
cttcattcac	ttaatctctt	caattggctt	tactggcttc	cacaattggt	tactgatgtt	10140
cgatataaac	caaattcgaa	ctttgttctg	attctctgca	agatggctgc	tgctcatcca	10200
cttcaagtat	tttaccacat	tcggggaggca	gttagcggtg	acgatattga	ctcggttctc	10260
gaagaagatt	acactgatga	gcaaattgtcg	atggatgttt	cggatgagga	ttgttttgca	10320
gacgatccac	catttgatag	aattctgaaa	atatgtctga	aatatcgctc	aactgatatt	10380
cgagtcttcc	atcgtgtcct	caaagaactt	gacgagatga	atgagacatg	ggttgaacgt	10440
cacttgcgct	atcgatctg	cctcaaggat	cagatgttca	aagatttctc	ggaacaaatg	10500
gacgcgacgt	tcaatgagat	gcaatattcg	gaggatgtga	ctatgatgac	gttgagatgg	10560
aggaaacagc	tggaaagaaga	cttgggtgtat	ttccaacaga	attataatct	tgatttcctg	10620
gagattcgta	acaagcgaaa	gatgatcgtg	acgaagggat	gtatgggagt	cgagaaaagt	10680
cagataatgt	tcgaaaaaga	gctgagtcaa	gtgttcacag	agccggccgg	catgcaagat	10740
gaatttgatt	ttgtcacaaa	tatgactaat	atgatggctc	cacagttgga	tattcatgca	10800
gtcgatgctc	cacgcccctca	gggatatat	cgtattgttc	tcgactggat	tcgagcgatt	10860
cgctcgctgt	tcgatcgact	tcacacgaaga	atccctctgg	aatcgccaag	cccatactct	10920
gccagattca	gccatcgta	aggatgcac	gaaatgccat	acgatttgct	caacgttttg	10980
cgcgccaaga	atcatactct	gatggcttcc	aatcaaacgg	ggcaatacat	atccatgctc	11040
tctcgatttg	agccaaactt	tgagattgtg	atcaaagggtg	gtcaagtgat	aagaaagatc	11100
tatatccgag	gacaaaaccgg	aaagagtgcg	gcgtttttatc	tgaagaaatc	tgtgcaggat	11160
gagccaacta	accgagttcc	acaaatgttc	aaacatcttg	atcacgttct	acaaaccgat	11220
agagagtcgg	cgagaagaca	tcttcatgct	ccaacagtgc	tgcatgatgag	agtcggacag	11280
aagacgacac	tctacgaagt	tgcatccgtt	caaccatattg	caatgccacc	ggattgtacc	11340
agaaactatc	cagcatcaca	aatcgacatt	gttcatccat	atgatgtgct	gactgccact	11400
ttcaatggaa	gttattatcc	ggatgatattg	gtattgcact	tctttgagag	attcgcccaa	11460
agttcttcat	ccatcggaca	acctcttcca	actccgacga	accaagatgg	aacagttgct	11520
ccgccacgac	taacggaagc	tcaccacatc	aagaatatta	tttatgaaga	ctttgcccga	11580
gatatgatcc	cattccgact	tctctacgac	tacctcactg	cacgatatcc	tgatccgggt	11640
atgtactatg	caatgaagaa	gcaattgctg	cacagtctcg	ccgtccctatc	cacaatcgaa	11700
tatcattgca	atctgacacc	aatgggacct	gatcaaatga	tgatgacaat	gaatactgga	11760
gtccttagca	atccttcata	tagattcgaa	atccgaggag	gacgatcact	tcatgatatt	11820
caacactttg	gacatgaagt	tccattccga	ttgactccaa	atctatcgat	tttggttggt	11880
gttgacacag	atggtgactt	gttatggagt	atggctgctg	cgtcaaaatg	tttgatgaag	11940
aaggaacctg	aagtatcat	gagaccgtta	gtatgggatg	aattcgccaa	caatacagat	12000
tgcgacaaat	cgcgtttgca	ggtattcgcg	tgtcatgcat	cgaattctta	catcaatggt	12060
gtcgcgagca	agcttcgaaa	cacgaatagc	gccgacgcca	aactcagaaa	ggacgattgt	12120
gtgtcgctga	tcagtcgagc	caaggattcg	gataatctgg	cccgaatgcc	accacactac	12180
cacgcgtggt	tctag					12195

<210> 14
 <211> 4064
 <212> PRT
 <213> *Caenorhabditis elegans*

<400> 14

Met	Asp	Pro	Ala	Met	Ala	Ser	Pro	Gly	Tyr	Arg	Ser	Val	Gln	Ser	Asp
1				5					10					15	
Arg	Ser	Asn	His	Leu	Thr	Glu	Leu	Glu	Thr	Arg	Ile	Gln	Asn	Leu	Ala
			20					25					30		
Asp	Asn	Ser	Gln	Arg	Asp	Asp	Val	Lys	Leu	Lys	Met	Leu	Gln	Glu	Ile
		35					40					45			
Trp	Ser	Thr	Ile	Glu	Asn	His	Phe	Thr	Leu	Ser	Ser	His	Glu	Lys	Val
	50					55					60				
Val	Glu	Arg	Leu	Ile	Leu	Ser	Phe	Leu	Gln	Val	Phe	Cys	Asn	Thr	Ser
65					70					75				80	
Pro	Gln	Phe	Ile	Ala	Glu	Asn	Asn	Thr	Gln	Gln	Leu	Arg	Lys	Leu	Met
				85					90					95	
Leu	Glu	Ile	Ile	Leu	Arg	Leu	Ser	Asn	Val	Glu	Ala	Met	Lys	His	His
			100					105					110		
Ser	Lys	Glu	Ile	Ile	Lys	Gln	Met	Met	Arg	Leu	Ile	Thr	Val	Glu	Asn
		115					120					125			
Glu	Glu	Asn	Ala	Asn	Leu	Ala	Ile	Lys	Ile	Val	Thr	Asp	Gln	Gly	Arg
		130				135					140				
Ser	Thr	Gly	Lys	Met	Gln	Tyr	Cys	Gly	Glu	Val	Ser	Gln	Ile	Met	Val
145					150					155				160	
Ser	Phe	Lys	Thr	Met	Val	Ile	Asp	Leu	Thr	Ala	Ser	Gly	Arg	Ala	Gly
			165						170					175	
Asp	Met	Phe	Asn	Ile	Lys	Glu	His	Lys	Ala	Pro	Pro	Ser	Thr	Ser	Ser
			180					185					190		
Asp	Glu	Gln	Val	Ile	Thr	Glu	Tyr	Leu	Lys	Thr	Cys	Tyr	Tyr	Gln	Gln
		195					200					205			
Thr	Val	Leu	Leu	Asn	Gly	Thr	Glu	Gly	Lys	Pro	Pro	Leu	Lys	Tyr	Asn
	210				215							220			
Met	Ile	Pro	Ser	Ala	His	Gln	Ser	Thr	Lys	Val	Leu	Leu	Glu	Val	Pro
225					230					235					240
Tyr	Leu	Val	Ile	Phe	Phe	Tyr	Gln	His	Phe	Lys	Thr	Ala	Ile	Gln	Thr
				245					250					255	
Glu	Ala	Leu	Asp	Phe	Met	Arg	Leu	Gly	Leu	Asp	Phe	Leu	Asn	Val	Arg
			260					265					270		
Val	Pro	Asp	Glu	Asp	Lys	Leu	Lys	Thr	Asn	Gln	Ile	Ile	Thr	Asp	Asp
		275					280					285			
Phe	Val	Ser	Ala	Gln	Ser	Arg	Phe	Leu	Ser	Phe	Val	Asn	Ile	Met	Ala
	290					295					300				
Lys	Ile	Pro	Ala	Phe	Met	Asp	Leu	Ile	Met	Gln	Asn	Gly	Pro	Leu	Leu
305					310					315				320	
Val	Ser	Gly	Thr	Met	Gln	Met	Leu	Glu	Arg	Cys	Pro	Ala	Asp	Leu	Ile
				325					330					335	
Ser	Val	Arg	Arg	Glu	Val	Leu	Met	Ala	Leu	Lys	Tyr	Phe	Thr	Ser	Gly
			340					345					350		
Glu	Met	Lys	Ser	Lys	Phe	Phe	Pro	Met	Leu	Pro	Arg	Leu	Ile	Ala	Glu
		355					360					365			
Glu	Val	Leu	Gly	Thr	Gly	Phe	Thr	Ala	Ile	Glu	His	Leu	Arg	Val	
	370				375					380					
Phe	Met	Tyr	Gln	Met	Leu	Ala	Asp	Leu	Leu	His	His	Met	Arg	Asn	Ser
385					390					395					400

Ile	Asp	Tyr	Glu	Met	Ile	Thr	His	Val	Ile	Phe	Val	Phe	Cys	Arg	Thr		
				405					410					415			
Leu	His	Asp	Pro	Asn	Asn	Ser	Ser	Gln	Val	Gln	Ile	Met	Ser	Ala	Arg		
			420					425					430				
Leu	Leu	Asn	Ser	Leu	Ala	Glu	Ser	Leu	Cys	Lys	Met	Asp	Ser	His	Asp		
		435					440					445					
Thr	Phe	Gln	Thr	Arg	Asp	Leu	Leu	Ile	Glu	Ile	Leu	Glu	Ser	His	Val		
	450					455					460						
Ala	Lys	Leu	Lys	Thr	Leu	Ala	Val	Tyr	His	Met	Pro	Ile	Leu	Phe	Gln		
465					470					475					480		
Gln	Tyr	Gly	Thr	Glu	Ile	Asp	Tyr	Glu	Tyr	Lys	Ser	Tyr	Glu	Arg	Asp		
				485					490					495			
Ala	Glu	Lys	Pro	Gly	Met	Asn	Ile	Pro	Lys	Asp	Thr	Ile	Arg	Gly	Val		
			500					505					510				
Pro	Lys	Arg	Arg	Ile	Arg	Arg	Leu	Ser	Ile	Asp	Ser	Val	Glu	Glu	Leu		
		515					520					525					
Glu	Phe	Leu	Ala	Ser	Glu	Pro	Ser	Thr	Ser	Glu	Asp	Ala	Asp	Glu	Ser		
	530					535					540						
Gly	Gly	Asp	Pro	Asn	Lys	Leu	Pro	Pro	Pro	Thr	Lys	Glu	Gly	Lys	Lys		
545					550					555					560		
Thr	Ser	Pro	Glu	Ala	Ile	Leu	Thr	Ala	Met	Ser	Thr	Met	Thr	Pro	Pro		
				565				570						575			
Pro	Leu	Ala	Ile	Val	Glu	Ala	Arg	Asn	Leu	Val	Lys	Tyr	Ile	Met	His		
			580					585					590				
Thr	Cys	Lys	Phe	Val	Thr	Gly	Gln	Leu	Arg	Ile	Ala	Arg	Pro	Ser	Gln		
	595					600					605						
Asp	Met	Tyr	His	Cys	Ser	Lys	Glu	Arg	Asp	Leu	Phe	Glu	Arg	Leu	Leu		
	610					615				620							
Arg	Tyr	Gly	Val	Met	Cys	Met	Asp	Val	Phe	Val	Leu	Pro	Thr	Thr	Arg		
625					630				635						640		
Asn	Gln	Pro	Gln	Met	His	Ser	Ser	Met	Arg	Thr	Lys	Asp	Glu	Lys	Asp		
				645					650					655			
Ala	Leu	Glu	Ser	Leu	Ala	Asn	Val	Phe	Thr	Thr	Ile	Asp	His	Ala	Ile		
			660					665					670				
Phe	Arg	Glu	Ile	Phe	Glu	Lys	Tyr	Met	Asp	Phe	Leu	Ile	Glu	Arg	Ile		
	675						680				685						
Tyr	Asn	Arg	Asn	Tyr	Pro	Leu	Gln	Leu	Met	Val	Asn	Thr	Phe	Leu	Val		
	690					695					700						
Arg	Asn	Glu	Val	Pro	Phe	Phe	Ala	Ser	Thr	Met	Leu	Ser	Phe	Leu	Met		
705					710					715					720		
Ser	Arg	Met	Lys	Leu	Leu	Glu	Val	Ser	Asn	Asp	Lys	Thr	Met	Leu	Tyr		
				725					730					735			
Val	Lys	Leu	Phe	Lys	Ile	Ile	Phe	Ser	Ala	Ile	Gly	Ala	Asn	Gly	Ser		
			740					745					750				
Gly	Leu	His	Gly	Asp	Lys	Met	Leu	Thr	Ser	Tyr	Leu	Pro	Glu	Ile	Leu		
	755						760					765					
Lys	Gln	Ser	Thr	Val	Leu	Ala	Leu	Thr	Ala	Arg	Glu	Pro	Leu	Asn	Tyr		
	770					775					780						
Phe	Leu	Leu	Leu	Arg	Ala	Leu	Phe	Arg	Ser	Ile	Gly	Gly	Gly	Ala	Gln		
785					790					795					800		
Asp	Ile	Leu	Tyr	Gly	Lys	Phe	Leu	Gln	Leu	Leu	Pro	Asn	Leu	Leu	Gln		
				805				810						815			
Phe	Leu	Asn	Lys	Leu	Thr	Asn	Leu	Gln	Ser	Cys	Gln	His	Arg	Ile	Gln		
		820						825					830				
Met	Arg	Glu	Leu	Phe	Val	Glu	Leu	Cys	Leu	Thr	Val	Pro	Val	Arg	Leu		
	835					840						845					
Ser	Ser	Leu	Leu	Pro	Tyr	Leu	Pro	Leu	Leu	Met	Asp	Pro	Leu	Val	Cys		
	850					855					860						

Ala Met Asn Gly Ser Pro Asn Ile Val Thr Gln Gly Leu Arg Thr Leu	865	870	875	880
Glu Leu Cys Val Asp Asn Leu Gln Pro Glu Tyr Leu Leu Glu Asn Met	885	890	895	
Leu Pro Val Arg Gly Ala Leu Met Gln Gly Leu Trp Arg Val Val Ser	900	905	910	
Lys Ala Pro Asp Thr Ser Ser Met Thr Ala Ala Phe Arg Ile Leu Gly	915	920	925	
Lys Phe Gly Gly Ala Asn Arg Lys Leu Leu Asn Gln Pro Gln Ile Leu	930	935	940	
Gln Val Ala Thr Leu Gly Asp Thr Val Gln Ser Tyr Ile Asn Met Glu	945	950	955	960
Phe Ser Arg Met Gly Leu Asp Gly Asn His Ser Ile His Leu Pro Leu	965	970	975	
Ser Glu Leu Met Arg Val Val Ala Asp Gln Met Arg Tyr Pro Ala Asp	980	985	990	
Met Ile Leu Asn Pro Ser Pro Ala Met Ile Pro Ser Thr His Met Lys	995	1000	1005	
Lys Trp Cys Met Glu Leu Ser Lys Ala Val Leu Leu Ala Gly Leu Gly	1010	1015	1020	
Ser Ser Gly Ser Pro Ile Thr Pro Ser Ala Asn Leu Pro Lys Ile Ile	1025	1030	1035	1040
Lys Lys Leu Leu Glu Asp Phe Asp Pro Asn Asn Arg Thr Thr Glu Val	1045	1050	1055	
Tyr Thr Cys Pro Arg Glu Ser Asp Arg Glu Leu Phe Val Asn Ala Leu	1060	1065	1070	
Leu Ala Met Ala Tyr Gly Ile Trp Asn Lys Asp Gly Phe Arg His Val	1075	1080	1085	
Tyr Ser Lys Phe Phe Ile Lys Val Leu Arg Gln Phe Ala Leu Ile Gly	1090	1095	1100	
Val Leu Glu Tyr Ile Gly Gly Asn Gly Trp Met Arg His Ala Glu Glu	1105	1110	1115	1120
Glu Gly Val Leu Pro Leu Cys Leu Asp Ser Ser Val Met Val Asp Ala	1125	1130	1135	
Leu Ile Ile Cys Leu Ser Glu Thr Ser Ser Ser Phe Ile Ile Ala Gly	1140	1145	1150	
Val Met Ser Leu Arg His Ile Asn Glu Thr Leu Ser Leu Thr Leu Pro	1155	1160	1165	
Asp Ile Asp Gln Met Ser Lys Val Pro Met Cys Lys Tyr Leu Met Glu	1170	1175	1180	
Lys Val Phe Lys Leu Cys His Gly Pro Ala Trp Tyr Ala Arg Ser Gly	1185	1190	1195	1200
Gly Ile Asn Ala Ile Gly Tyr Met Ile Glu Ser Phe Pro Arg Lys Phe	1205	1210	1215	
Val Met Asp Phe Val Ile Asp Val Val Asp Ser Ile Met Glu Val Ile	1220	1225	1230	
Leu Gly Thr Val Glu Glu Ile Ser Ser Gly Ser Ala Asp Ser Ala Tyr	1235	1240	1245	
Asp Cys Leu Lys Lys Met Met Arg Val Tyr Phe Ile Lys Glu Glu Gly	1250	1255	1260	
Gln Glu Glu Glu Asn Leu Thr Leu Ala Thr Ile Phe Val Ser Ala Ile	1265	1270	1275	1280
Ser Lys His Tyr Phe His Ser Asn Glu Arg Val Arg Glu Phe Ala Ile	1285	1290	1295	
Gly Leu Met Asp His Cys Met Val His Ser Arg Leu Ala Pro Ser Leu	1300	1305	1310	
Asp Lys Phe Tyr Tyr Arg Phe Lys Glu Phe Phe Glu Pro Glu Leu Met	1315	1320	1325	

Arg Val Leu Thr Thr Val Pro Thr Met Ser Leu Ala Asp Ala Gly Gly
 1330 1335 1340
 Ser Leu Asp Gly Val Gln Asn Tyr Met Phe Asn Cys Pro Asp Gly Phe
 1345 1350 1355 1360
 Asp Phe Glu Lys Asp Met Asp Met Tyr Lys Arg Tyr Leu Ser His Leu
 1365 1370 1375
 Leu Asp Ile Ala Gln Thr Asp Thr Phe Thr Leu Asn Gln Arg Asn Ala
 1380 1385 1390
 Phe Lys Lys Cys Glu Thr Cys Pro Ser His Phe Leu Pro Pro Phe Pro
 1395 1400 1405
 Ile Thr Thr His Ile Asp Ser Met Arg Ala Ser Ala Leu Gln Cys Leu
 1410 1415 1420
 Val Ile Ala Tyr Asp Arg Met Lys Lys Gln Tyr Ile Asp Lys Gly Ile
 1425 1430 1435 1440
 Glu Leu Gly Asp Glu His Lys Met Ile Glu Ile Leu Ala Leu Arg Ser
 1445 1450 1455
 Ser Lys Ile Thr Val Asp Gln Val Tyr Glu Ser Asp Glu Ser Trp Arg
 1460 1465 1470
 Arg Leu Met Thr Val Leu Leu Arg Ala Val Thr Asp Arg Glu Thr Pro
 1475 1480 1485
 Glu Ile Ala Glu Lys Leu His Pro Ser Leu Leu Lys Val Ser Pro Ile
 1490 1495 1500
 Ser Thr Ile Ile Ile Ala Thr Phe Gly Ala Ser Tyr Ile Arg Asn Ile
 1505 1510 1515 1520
 Ser Gly Ala Gly Asp Asp Ser Asp Ser Asp Arg His Ile Ser Tyr Asn
 1525 1530 1535
 Asp Ile Met Lys Phe Lys Cys Leu Val Glu Leu Asn Pro Lys Ile Leu
 1540 1545 1550
 Val Thr Lys Met Ala Val Asn Leu Ala Asn Gln Met Val Lys Tyr Lys
 1555 1560 1565
 Met Ser Asp Lys Ile Ser Arg Ile Leu Ser Val Pro Ser Ser Phe Thr
 1570 1575 1580
 Glu Glu Glu Leu Asp Asp Phe Glu Ala Glu Lys Met Lys Gly Ile Arg
 1585 1590 1595 1600
 Glu Leu Asp Met Ile Gly His Thr Val Lys Met Leu Ala Gly Cys Pro
 1605 1610 1615
 Val Thr Thr Phe Thr Glu Gln Ile Ile Val Asp Ile Ser Arg Phe Ala
 1620 1625 1630
 Ala His Phe Glu Tyr Ala Tyr Ser Gln Asp Val Leu Val Asn Trp Ile
 1635 1640 1645
 Asp Asp Val Thr Val Ile Leu Asn Lys Ser Pro Lys Asp Val Trp Lys
 1650 1655 1660
 Phe Phe Leu Ser Arg Glu Ser Ile Leu Asp Pro Ala Arg Arg Ser Phe
 1665 1670 1675 1680
 Ile Arg Arg Ile Ile Val Tyr Gln Ser Ser Gly Pro Leu Arg Gln Glu
 1685 1690 1695
 Phe Met Asp Thr Pro Glu Tyr Phe Glu Lys Leu Ile Asp Leu Asp Asp
 1700 1705 1710
 Glu Glu Asn Lys Asp Glu Asp Glu Arg Lys Ile Trp Asp Arg Asp Met
 1715 1720 1725
 Phe Ala Phe Ser Ile Val Asp Arg Ile Ser Lys Ser Cys Pro Glu Trp
 1730 1735 1740
 -Leu Ile Ser Pro Asn Ser Pro Ile Pro Arg Ile Lys Lys Leu Phe Ser
 1745 1750 1755 1760
 Glu Thr Glu Phe Asn Glu Arg Tyr Val Val Arg Ala Leu Thr Glu Val
 1765 1770 1775
 Lys Lys Phe Gln Glu Glu Ile Ile Val Lys Arg Met Thr Glu His Lys
 1780 1785 1790

Tyr Lys Val Pro Lys Leu Ile Leu Asn Thr Phe Leu Arg Tyr Leu Arg
 1795 1800 1805
 Leu Asn Ile Tyr Asp Tyr Asp Leu Phe Ile Val Ile Ala Ser Cys Phe
 1810 1815 1820
 Asn Gly Asn Phe Val Thr Asp Leu Ser Phe Leu Arg Glu Tyr Leu Glu
 1825 1830 1835 1840
 Thr Glu Val Ile Pro Lys Val Pro Leu Gln Trp Arg Arg Glu Leu Phe
 1845 1850 1855
 Leu Arg Ile Met Gln Lys Phe Asp Thr Asp Pro Gln Thr Ala Gly Thr
 1860 1865 1870
 Ser Met Gln His Val Lys Ala Leu Gln Tyr Leu Val Ile Pro Thr Leu
 1875 1880 1885
 His Trp Ala Phe Glu Arg Tyr Asp Thr Asp Glu Ile Val Gly Thr Ala
 1890 1895 1900
 Pro Ile Asp Asp Ser Asp Ser Ser Met Asp Val Asp Pro Ala Gly Ser
 1905 1910 1915 1920
 Ser Asp Asn Leu Val Ala Arg Leu Thr Ser Val Ile Asp Ser His Arg
 1925 1930 1935
 Asn Tyr Leu Ser Asp Gly Met Val Ile Val Phe Tyr Gln Leu Cys Thr
 1940 1945 1950
 Leu Phe Val Gln Asn Ala Ser Glu His Ile His Asn Asn Asn Cys Lys
 1955 1960 1965
 Lys Gln Gly Gly Arg Leu Arg Ile Leu Met Leu Phe Ala Trp Pro Cys
 1970 1975 1980
 Leu Thr Met Tyr Asn His Gln Asp Pro Thr Met Arg Tyr Thr Gly Phe
 1985 1990 1995 2000
 Phe Phe Leu Ala Asn Ile Ile Glu Arg Phe Thr Ile Asn Arg Lys Ile
 2005 2010 2015
 Val Leu Gln Val Phe His Gln Leu Met Thr Thr Tyr Gln Gln Asp Thr
 2020 2025 2030
 Arg Asp Gln Ile Arg Lys Ala Ile Asp Ile Leu Thr Pro Ala Leu Arg
 2035 2040 2045
 Thr Arg Met Glu Asp Gly His Leu Gln Ile Leu Ser His Val Lys Lys
 2050 2055 2060
 Ile Leu Ile Glu Glu Cys His Asn Leu Gln His Val Gln His Val Phe
 2065 2070 2075 2080
 Gln Met Val Val Arg Asn Tyr Arg Val Tyr Tyr His Val Arg Leu Glu
 2085 2090 2095
 Leu Leu Thr Pro Leu Leu Asn Gly Val Gln Arg Ala Leu Val Met Pro
 2100 2105 2110
 Asn Ser Val Leu Glu Lys Phe Ser Trp Gln Thr Arg Arg His Ala Val
 2115 2120 2125
 Glu Ile Cys Glu Met Val Ile Lys Trp Glu Leu Phe Arg Thr Leu Lys
 2130 2135 2140
 Thr Asp His Ile Ile Ser Asp Glu Glu Ala Leu Glu Val Asp Lys Gln
 2145 2150 2155 2160
 Leu Asp Lys Leu Arg Thr Ala Ser Ser Thr Asp Arg Phe Asp Phe Glu
 2165 2170 2175
 Glu Ala His Asn Lys Arg Asp Met Pro Asp Ala Gln Arg Thr Ile Ile
 2180 2185 2190
 Lys Glu His Ala Asp Val Ile Val Asn Met Leu Val Arg Phe Cys Met
 2195 2200 2205
 Thr Phe His Gln Asn Ser Gly Ser Ser Ser Thr Ser Gln Ser Gly Asn
 2210 2215 2220
 His Gly Val Glu Leu Thr Lys Lys Cys Gln Leu Leu Leu Arg Ala Ala
 2225 2230 2235 2240
 Leu Arg Pro Ser Met Trp Gly Glu Phe Val Ser Phe Arg Leu Thr Met
 2245 2250 2255

Ile Glu Lys Phe Leu Ser Ile Pro Asn Asp Asn Ala Leu Arg Asn Asp
 2260 2265 2270
 Ile Ser Ser Thr Ala Tyr Ala Asn Thr Ile Gln Asn Ala Gln His Thr
 2275 2280 2285
 Leu Asp Met Leu Cys Asn Ile Ile Pro Val Met Pro Lys Thr Ser Leu
 2290 2295 2300
 Met Thr Met Met Arg Gln Leu Gln Arg Pro Leu Ile Gln Cys Leu Asn
 2305 2310 2315 2320
 Asn Gly Ala Gln Asn Phe Lys Met Thr Arg Leu Val Thr Gln Ile Val
 2325 2330 2335
 Ser Arg Leu Leu Glu Lys Thr Asn Val Ser Val Asn Gly Leu Asp Glu
 2340 2345 2350
 Leu Glu Gln Leu Asn Gln Tyr Ile Ser Arg Phe Leu His Glu His Phe
 2355 2360 2365
 Gly Ser Leu Leu Asn Cys Arg Asn Leu Ser Gly Pro Val Leu Gly Val
 2370 2375 2380
 Leu Gly Ala Phe Ser Leu Leu Arg Thr Ile Cys Gly His Glu Pro Ala
 2385 2390 2395 2400
 Tyr Leu Asp His Leu Met Pro Ser Phe Val Lys Val Met Glu Arg Ala
 2405 2410 2415
 Ala Lys Glu His Leu Ala Tyr Val Ala Asn Ser Gln Asp Gly Asn Met
 2420 2425 2430
 Val Lys Asn Phe Phe Pro Asp Val Ala Glu Leu Leu Cys Ala Cys Met
 2435 2440 2445
 Glu Leu Val Arg Pro Arg Val Asp His Ile Ser Met Glu Ile Lys Arg
 2450 2455 2460
 Ser Ile Val Gly Gly Ile Ile Ala Glu Leu Ile Ile Lys Ser Asn His
 2465 2470 2475 2480
 Asp Lys Ile Ile Gln Thr Ser Val Lys Leu Leu Gly Ala Met Ile Ser
 2485 2490 2495
 Thr Gln Asp Met Glu Phe Thr Ile Leu Thr Val Leu Pro Leu Leu Val
 2500 2505 2510
 Arg Ile Gln Ser Ile Ile Val Thr Lys Phe Lys Asn Cys Lys Asp Leu
 2515 2520 2525
 Ile Ala Asp Tyr Leu Val Val Val Ile Thr Val Phe Glu Asn Ser Glu
 2530 2535 2540
 Tyr Arg Asn Ser Glu Ala Gly Ser Arg Leu Trp Glu Gly Phe Phe Trp
 2545 2550 2555 2560
 Gly Leu Lys Ser Ser Asp Pro Gln Thr Arg Glu Lys Phe Ser Ile Val
 2565 2570 2575
 Trp Glu Lys Thr Trp Pro His Met Ala Thr Val Asp Ile Ala His Arg
 2580 2585 2590
 Met Lys Tyr Ile Met Gln Asn Gln Asp Trp Ser Lys Phe Lys His Ala
 2595 2600 2605
 Phe Trp Leu Lys Phe Ala Leu Trp Gly Met Leu Arg Thr Ile Ala Lys
 2610 2615 2620
 Arg Pro Thr Asp Pro Asn Asn Lys Arg Lys Lys Val Ile Leu Leu Asn
 2625 2630 2635 2640
 Cys Ala Thr Pro Trp Arg Thr Ile Glu Tyr Ala Ala Lys Leu Lys Asp
 2645 2650 2655
 Gln Pro Met Glu Val Glu Thr Glu Met Lys Arg Glu Glu Pro Glu Pro
 2660 2665 2670
 Met Glu Val Asp Glu Lys Asp Ser Gln Asp Asp Ser Lys Asp Ala Gly
 2675 2680 2685
 Glu Pro Lys Glu Lys Glu Lys Leu Thr Leu Glu Leu Leu Ala Gly
 2690 2695 2700
 Gln Gln Glu Leu Leu Asp Glu Ala Ser Asn Tyr Asp Phe Ala Asp Ala
 2705 2710 2715 2720

Leu Asp Thr Val Ser Gln Ile Thr Phe Ala Leu Asn Glu Asn Gln Val
 2725 2730 2735
 Thr Ser Lys Met Trp Val Val Leu Phe Lys Ser Phe Trp Ser Ser Leu
 2740 2745 2750
 Ser Gln Ser Glu Ile Glu Asp Phe Thr Ala Leu Val Val Pro Phe Met
 2755 2760 2765
 Ser Ser Gly Val His Asn Asn Tyr Gln Thr Gly Val Gln Asp Ser Val
 2770 2775 2780
 Leu Ala Val Trp Leu Glu Ala Val Gly Asp Ala Val His Leu Pro Ser
 2785 2790 2795 2800
 Arg Leu Ile Glu Phe Ile Ser Ser Lys His Glu Cys Trp His Thr Gly
 2805 2810 2815
 Ile Arg Leu Leu Glu Asn His Ile Trp Thr Ile Pro Lys Gln Leu Asn
 2820 2825 2830
 Asn Thr Leu Leu Arg Glu Met Lys Val Ala Pro Gly Leu Ala Gly Asp
 2835 2840 2845
 Ile Glu Thr Leu Glu Ser Leu Gly Thr Leu Tyr Asn Glu Ile Ser Glu
 2850 2855 2860
 Phe Asp Gln Phe Ala Ala Ile Trp Glu Arg Arg Ala Val Phe Pro Asp
 2865 2870 2875 2880
 Thr Met Arg Ala Met Ser Ala Met Gln Leu Gly Asp Met Glu Leu Ala
 2885 2890 2895
 Gln Ser Tyr Leu Glu Lys Ser Met Ser Ser Thr Tyr Glu Thr Leu Ala
 2900 2905 2910
 Pro Thr Ile Asn Pro Asn Asn Thr Ser Asn Ser Glu Lys His Val Ser
 2915 2920 2925
 Pro Ile Ile Asp Lys Glu Tyr Asp His Trp Met Glu Met Tyr Ile Thr
 2930 2935 2940
 Asn Cys Ser Glu Leu Leu Gln Trp Gln Asn Val Ala Asp Val Cys Asn
 2945 2950 2955 2960
 Gly Lys Asp Met Gln His Val Arg Gly Leu Ile Asn Ala Ala Ser His
 2965 2970 2975
 Ile Pro Asp Trp Asn Val Val Glu Glu Cys Lys Ser Gln Ile Ala Gly
 2980 2985 2990
 Cys Ile Pro Pro Ser Phe His Leu Asp Tyr Thr Leu Phe Asn Leu Met
 2995 3000 3005
 Ser Thr Val Met Arg Met Asn Glu Asn Ser Ser Pro Thr His Met Lys
 3010 3015 3020
 Glu Arg Cys Lys Ile Ala Ile Gln Glu Cys Thr Glu Ala His Ile Ser
 3025 3030 3035 3040
 Arg Trp Arg Ala Leu Pro Ser Val Val Ser Tyr Gly His Val Lys Ile
 3045 3050 3055
 Leu Gln Ala Met Asn Leu Val Arg Glu Ile Glu Glu Ser Thr Asp Ile
 3060 3065 3070
 Arg Ile Ala Leu Leu Glu Ala Pro Ser Asn Lys Val Asp Gln Ala Leu
 3075 3080 3085
 Met Gly Asp Met Lys Ser Leu Met Lys Val Phe Arg Asn Arg Thr Pro
 3090 3095 3100
 Thr Thr Ser Asp Asp Met Gly Phe Val Ser Thr Trp Tyr Asp Trp Arg
 3105 3110 3115 3120
 Asn Gln Ile His Gly Met Met Leu Gln Arg Phe Glu Tyr Trp Asp Lys
 3125 3130 3135
 Val Gly Leu Asn Val Ala Ala Thr Gly Asn Gln Ser Ile Val Pro Ile
 3140 3145 3150
 His Ser Met Ala Gln Ala Gln Leu Ala Val Ala Lys His Ala Lys Asn
 3155 3160 3165
 Leu Gly Phe His Asn Leu Thr Lys Asp Leu Leu Asn Lys Leu Ala Gly
 3170 3175 3180

Leu Thr Ala Ile Pro Met Met Asp Ala Gln Asp Lys Val Cys Thr Tyr
 3185 3190 3195 3200
 Gly Lys Thr Leu Arg Asp Met Ala Asn Ser Ala Ala Asp Glu Arg Val
 3205 3210 3215
 Lys Asn Glu Leu Leu Cys Glu Ala Leu Glu Val Leu Glu Asp Val Arg
 3220 3225 3230
 Ile Asp Asp Leu Gln Lys Asp Gln Val Ala Ala Leu Leu Tyr His Arg
 3235 3240 3245
 Ala Asn Ile His Ser Val Leu Asp Gln Ala Glu Asn Ala Asp Tyr Thr
 3250 3255 3260
 Phe Ser Ala Ala Ser Gln Leu Val Asp Leu Gln Asn Ser Val Thr Thr
 3265 3270 3275 3280
 Thr Gly Ile Lys Leu Met Lys Asn Trp Gly His His Leu Tyr Lys Arg
 3285 3290 3295
 Phe Phe Ser Thr Thr Val Cys Lys Glu Thr Gly Asn Asn Phe Gly Arg
 3300 3305 3310
 Gln Ala Leu Ala Cys Tyr Phe Ile Ala Ala Arg Val Asp Asn Asp Ile
 3315 3320 3325
 Lys Ala Arg Lys Pro Ile Ala Lys Ile Leu Trp Leu Ser Lys His Leu
 3330 3335 3340
 Asn Ala Cys Gly Ser His Glu Val Met Asn Arg Val Ile Lys Lys Gln
 3345 3350 3355 3360
 Leu His Ser Leu Asn Leu Phe Asn Trp Leu Tyr Trp Leu Pro Gln Leu
 3365 3370 3375
 Val Thr Asp Val Arg Tyr Lys Pro Asn Ser Asn Phe Val Leu Ile Leu
 3380 3385 3390
 Cys Lys Met Ala Ala Ala His Pro Leu Gln Val Phe Tyr His Ile Arg
 3395 3400 3405
 Glu Ala Val Ser Val Asp Asp Ile Asp Ser Val Leu Glu Glu Asp Tyr
 3410 3415 3420
 Thr Asp Glu Gln Met Ser Met Asp Val Ser Asp Glu Asp Cys Phe Ala
 3425 3430 3435 3440
 Asp Asp Pro Pro Phe Asp Arg Ile Leu Lys Ile Cys Leu Lys Tyr Arg
 3445 3450 3455
 Pro Thr Asp Ile Arg Val Phe His Arg Val Leu Lys Glu Leu Asp Glu
 3460 3465 3470
 Met Asn Glu Thr Trp Val Glu Arg His Leu Arg His Ala Ile Cys Leu
 3475 3480 3485
 Lys Asp Gln Met Phe Lys Asp Phe Ser Glu Gln Met Asp Ala Thr Phe
 3490 3495 3500
 Asn Glu Met Gln Tyr Ser Glu Asp Val Thr Met Met Thr Leu Arg Trp
 3505 3510 3515 3520
 Arg Lys Gln Leu Glu Glu Asp Leu Val Tyr Phe Gln Gln Asn Tyr Asn
 3525 3530 3535
 Leu Asp Phe Leu Glu Ile Arg Asn Lys Arg Lys Met Ile Val Thr Lys
 3540 3545 3550
 Gly Cys Met Gly Val Glu Lys Ser Gln Ile Met Phe Glu Lys Glu Leu
 3555 3560 3565
 Ser Gln Val Phe Thr Glu Pro Ala Gly Met Gln Asp Glu Phe Asp Phe
 3570 3575 3580
 Val Thr Asn Met Thr Asn Met Met Val Ser Gln Leu Asp Ile His Ala
 3585 3590 3595 3600
 Val Asp Ala Pro Arg Pro Gln Gly Tyr Ile Arg Ile Val Leu Asp Trp
 3605 3610 3615
 Ile Arg Ala Ile Arg Arg Arg Phe Asp Arg Leu Pro Arg Arg Ile Pro
 3620 3625 3630
 Leu Glu Ser Ser Ser Pro Tyr Leu Ala Arg Phe Ser His Arg Thr Gly
 3635 3640 3645

Cys Ile Glu Met Pro Tyr Asp Leu Leu Asn Val Leu Arg Ala Lys Asn
 3650 3655 3660
 His Thr Leu Met Ala Ser Asn Gln Thr Gly Gln Tyr Ile Ser Met Leu
 3665 3670 3675 3680
 Ser Arg Phe Glu Pro Asn Phe Glu Ile Val Ile Lys Gly Gly Gln Val
 3685 3690 3695
 Ile Arg Lys Ile Tyr Ile Arg Gly Gln Thr Gly Lys Ser Ala Ala Phe
 3700 3705 3710
 Tyr Leu Lys Lys Ser Val Gln Asp Glu Pro Thr Asn Arg Val Pro Gln
 3715 3720 3725
 Met Phe Lys His Leu Asp His Val Leu Gln Thr Asp Arg Glu Ser Ala
 3730 3735 3740
 Arg Arg His Leu His Ala Pro Thr Val Leu Gln Met Arg Val Gly Gln
 3745 3750 3755 3760
 Lys Thr Thr Leu Tyr Glu Val Ala Ser Val Gln Pro Tyr Ala Met Pro
 3765 3770 3775
 Pro Asp Cys Thr Arg Asn Tyr Pro Ala Ser Gln Ile Asp Ile Val His
 3780 3785 3790
 Pro Tyr Asp Val Leu Thr Ala Thr Phe Asn Gly Ser Tyr Tyr Pro Asp
 3795 3800 3805
 Asp Met Val Leu His Phe Phe Glu Arg Phe Ala Gln Ser Ser Ser Ser
 3810 3815 3820
 Ile Gly Gln Pro Leu Pro Thr Pro Thr Asn Gln Asp Gly Thr Val Ala
 3825 3830 3835 3840
 Pro Pro Arg Leu Thr Glu Ala His His Ile Lys Asn Ile Ile Tyr Glu
 3845 3850 3855
 Asp Phe Ala Arg Asp Met Ile Pro Phe Arg Leu Leu Tyr Asp Tyr Leu
 3860 3865 3870
 Thr Ala Arg Tyr Pro Asp Pro Val Met Tyr Tyr Ala Met Lys Lys Gln
 3875 3880 3885
 Leu Leu His Ser Leu Ala Val Leu Ser Thr Ile Glu Tyr His Cys Asn
 3890 3895 3900
 Leu Thr Pro Met Gly Pro Asp Gln Met Met Met Thr Met Asn Thr Gly
 3905 3910 3915 3920
 Val Leu Ser Asn Pro Ser Tyr Arg Phe Glu Ile Arg Gly Gly Arg Ser
 3925 3930 3935
 Leu His Asp Ile Gln His Phe Gly His Glu Val Pro Phe Arg Leu Thr
 3940 3945 3950
 Pro Asn Leu Ser Ile Leu Val Gly Val Ala Gln Asp Gly Asp Leu Leu
 3955 3960 3965
 Trp Ser Met Ala Ala Ala Ser Lys Cys Leu Met Lys Lys Glu Pro Glu
 3970 3975 3980
 Val Ile Met Arg Pro Leu Val Trp Asp Glu Phe Ala Asn Asn Thr Asp
 3985 3990 3995 4000
 Cys Asp Lys Ser Arg Leu Gln Val Phe Ala Cys His Ala Ser Asn Ser
 4005 4010 4015
 Tyr Ile Asn Gly Val Ala Ser Lys Leu Arg Asn Thr Asn Ser Ala Asp
 4020 4025 4030
 Ala Lys Leu Arg Lys Asp Asp Cys Val Ser Leu Ile Ser Arg Ala Lys
 4035 4040 4045
 Asp Ser Asp Asn Leu Ala Arg Met Pro Pro Thr Tyr His Ala Trp Phe
 4050 4055 4060

<210> 15

<211> 4896

<212> DNA

<213> Caenorhabditis elegans

<400> 15

```

ttgttttcgg attttttgtg tgcttcgtag ttgctccgat gatgccggat tcaacatttg 60
aatgtaacat ttgaattttg aaattgaagg aattcatttg aatctaaagc ttgcagggtc 120
aagaccgata cattcttgca acacatgact cgaaagtatg taggaaaaat tgaagttgga 180
aacttggaat ttgatgaaaa agtacagtaa tccattctct cttatttcgc aactttcttc 240
gatttttgat ttttcctaga ttttttaagc taaaattttg ctgttttatt ttcatttttc 300
atgcttttca atttcggttt tcaacaaaat tatgtttttc agagaaaatc tcgtgaacaa 360
taactcggct actgtaccat ttaaaggcgc acaccttttc gcgcagcatt gatttaaatt 420
tttttgctcg tggctcaaca gtgcaatgga catctagata tctgaaattt taccactgaa 480
ttcagttcat tttttaagca tcttcaaaaa ttgctgtttt cctaattttc ttgtgatcgt 540
tttttttttg aaagtacaat cgtacattat aaataactat ttttcaattc gaataattta 600
attcaagatc atttcgcaa ataatgcct tgaaacgtta tgccgcggtc aattttcaac 660
cacccttggt attctttttt gaattgccgc cctttttccc tgtggccggc gcagtgcggc 720
cgaggttggg ttctaggcca gccggcgcgt tttatttttt tcgagcatga tttcacaatt 780
atttcttgca tttttaaggt tttttattga taaaatagta aaactaacia cggataatat 840
tattttaaaa ttaaaaaact agtttggttca tttttggatc gattttttaga tgttggtcat 900
ggattatgca cgcaagaaag tactatcgtt cacatttgat tgctatatta ttgaatattg 960
aatttttcac acaaaattgt actattttcc gatattttatc atgaccgagc cgaagaagga 1020
gattatagag gacgaaaatc atggaatata caagaaaata ccaacagatc ccaggcaata 1080
cgagaaagt ttacagggat gccggttatt ggtcatgatg gcttcacaag aagaagaaag 1140
ttagttttta catctattta aacacatttt ccaattattt tcaggatggg ccgaagttat 1200
ttcaagatgc cgagctgcaa atggttcaat taaattctat gtccattata tcgattgcaa 1260
ccgaagactt gacgaatggg ttcatgtctga taggctcaat tttagcgtcgt gtgagctacc 1320
aaaaaaagga ggaaagaaag gagcacactt gcgggaagaa aagtgagaaa tctataaact 1380
tttcaaaaga ttttaaatag ttttatcaat tcataattat ttcagtcgag attcgaatga 1440
aaatgaagga aagaaaagcg gccgaaaacg aaagattcca ctacttccga tggatgatct 1500
caaggcggaa tccgtagatc cattacaagc aatttcaacg atgaccagcg gatctactcc 1560
aagtcttcga ggttccatgt cgatggtcgg ccatagttaa gatgcaatga caaggatccg 1620
aaatgtcgaa tgcattgaac taggaagatc acgaattcag ccatggtact ttgcacctta 1680
tccacaacaa ttgacaagtt tggattgtat ttattttgct gaattttgtc tgaaatatct 1740
aaagtcgaaa acttgtctga aacggcacat ggtgagtgtt tcgagttata gaaaatgacc 1800
gaatataaat aactgttttc aaaattcaaa aattttcaat tttccaaaaa tgaaagaatc 1860
ggtgaattcg aaaaaattcg agttcttggt tgtttttggc tgaatttttc ggtttttctt 1920
gctttttccg ttgatattag ttttgaaaca atgtttttta aattttccgg catcgaaaaa 1980
aatcgcaaat tctgggaatt tgctccaaaa attgcatttt tgaaatactt ttttgcgaaa 2040
acgaaaaaaa aattcacaaa cgggtgtttca aaccaaattt atcgtaatca aaaaagtttc 2100
gcaaataggc cattattctg cgtgggaatt caaattaaaa tcagctactt tttctatttt 2160
gcaaatagga aaaaaaacgt aaaaaataga caaattttta atttttttaa ttaattacat 2220
cgggtccatac tcttcatttt ctatcattta attaaaatgc ccaattctaa ttaattttat 2280
ttcaggaaaa atgtgcaatg tgtcacccac ctggcaatca aatctacagt cacgataaac 2340
tttcatTTTT tgaaatcgac ggccgcaaaa acaaaagcta tgcgcagaat ctatgcctgc 2400
ttgccaaact ttttctggat cacaagactc tttactatga cacggatcca tttttgttct 2460
atgtgctaac cgaagaagac gagaagggtc atcatatagt tggatacttt tcaaaagaaa 2520
aagaatcagc tgaagaatat aatggttgcgt gtattcttgt gttacctcca tttcaaaaga 2580
aaggatacgg aagtttgctc atcgaattca gctatgaact ctogaaaatt gaacagaaga 2640
caggatcacc cgaaaaacca ctatcagatt tgggacttct ctcatatcga tcgtactggg 2700
caatggccat catgaaagag ctttttcgcat tcaaaagacg acatccaggc gaagatatca 2760
cagttcagga catttcacaa agtacatcga ttaaacgaga agatgttgtg tcaacgttac 2820
agcaacttga tctatacaaa tactataagg gatcatacat aattgtgatt agtgatgaaa 2880
agcgtcaagt ttatgagaaa cggattgagg ctgcgaaaaa gaagacacga attaatccag 2940
cagctctgca atggcgaccc aaagagtacg gaaagaaaag agtgagtttt tttcaatcaa 3000
aaattcgtgt ttacggctaa aaactgaaaa ttaaaattaa attaaattcg tgataacatt 3060
tttttttcaa aaaaccaaaa aaaaacaatt tcgttttttg cagaaccaaa aaaaaattt 3120
aaaaaaaaac ggtttacgcc ctatttcata caaacaacag aaattgcact tttttgagca 3180
aatttgaccc tacaattttt ttccagtttt ttgctctttt tcaaaaaaaa acacctaaac 3240
actggaaata ctaaatacta aggaaaaaaa tggaaatact ggtttacagt gtcaaaaaat 3300
tgaaattttc taataaaatc atttttcttt ttactaaatt tatcaaaaat ttataactca 3360
aatctttcag tttttgcgaa ttttttttcg aaaaaacgaa aaaaaataaa cctaatttta 3420

```

```

accaaattgt aattttgaaa aatctggaac gtccggaaaa ctgaaaaatt aaaaaaaaaa 3480
cttttcagaa atttatTTTT aaaaaaccgt ttttttaaat caaattttgt atatgttgat 3540
gagaaaaaaa aatagaaatc aatgttttta agtttttaaaa gaaaaattta ttttaattat 3600
tttagtttta ataaggtatt taaacagtaa caaggatgtc ggtttttcga ttttcgaaa 3660
aactaaaaaa ttgtctTTTT cgatttttta atcgaaaaaa aatagaaata ttttcacaaa 3720
acatactatt ctcttaaaaa aaagaatagt ggcagatttt aaataatttt tgaactctcg 3780
caattttttt cgaaatatcc aaaaatcgaa aaaccggcac aaaagcaaaa agtctccggg 3840
aatatatctt taaattattt tatgaacttt tttttcaggc gcagatcatg ttctagcaac 3900
aacgacatgt gttctcgcca cgacgatctc aacctgtaca ttaaaatata acactccgtt 3960
ttatctcgca tctacacacc gaaaagctta cgctatccct ttatcattcc cacaccgtc 4020
agagagcgta cgcctcattt catttcattt gttctgtgta ataatttgac ttattagtca 4080
cttatttttt taatgaaatt attcttgaat ttcataatct tcttggtgca gttcaaataa 4140
ttaaaattca tcatatagac aagtaagttt ataactgcaa aagtgaagtt ttctaatacat 4200
taagcgttct gaagatatcc ggcaaccgcc tgagcgatca gatcacggcg ggaacgagtt 4260
gaggcgtaga catgcttgca gccagtgaca acctgaaaga tattcaaaaa attaatttca 4320
ggactcgaat ttttaacaat ctgaataaaa aaatccaaaa ttgtatatata tagagttttt 4380
tgaaatctaa gcgaaagcgc gctccaatgt aaaacgaaaa gtgctccgcc cctaaacgtt 4440
gggtcccggt aggaatttgt tattttttcg gttattttctg actatattat aatttcgaaa 4500
cgacaagtat tttaaacatc atttcgacat aaaaaatatg taaaacaaca aaaaacaate 4560
gaaaaaatag tgaaaaagtt tgaatttaca gtctcgccgc ctctaccga gacctaacgt 4620
taggaggcgg agcgttttcc tttggcattg aagcgcgctt gctgcggccc cataattaat 4680
aacttacagc ctttgcaaaag tccttcttct gttcatctc aatctcgtca atgtattgat 4740
tggacaactt ctcaatctcg gactgttccg cattttcatc cttcaatttt ttgtattgag 4800
ccttgaattg agccaccttc tcctctccga aagccttaac cgaataactcc ttacaagctt 4860
ctttcaactt gccctcggcc ttctccttgg catctc 4896

```

<210> 16

<211> 1377

<212> DNA

<213> *Caenorhabditis elegans*

<400> 16

```

atgaccgagc cgaagaagga gattatagag gacgaaaatc atggaatata caagaaaata 60
ccaacagatc ccaggcaata cgagaaagtt acagagggat gccggttatt ggtcatgatg 120
gcttcacaag aagaagaaaag atgggcccga gttattttcaa gatgccgagc tgcaaatggg 180
tcaattaaat tctatgtcca ttatatcgat tgcaaccgaa gacttgacga atgggttcag 240
tctgataggc tcaatttagc gtctgtgtgag ctacaaaaaa aaggaggaaa gaaaggagca 300
cacttgccgg aagaaaatcg agattcgaat gaaaatgaag gaaagaaaag cggccgaaaa 360
cgaaagattc cactacttcc gatggatgat ctcaaggcgg aatccgtaga tccattacaa 420
gcaatttcaa cgatgaccag cggatctact ccaagtcttc gaggttccat gtcgatggtc 480
ggccatagtg aagatgcaat gacaaggatc cgaaatgtcg aatgcattga actaggaaga 540
tcacgaattc agccatggta ctttgcacct tatccacaac aattgacaag tttggattgt 600
atttatattt gcgaattttg tctgaaatat cttaaagtcga aaacttgtct gaaacggcac 660
atggaaaaat gtgcaatgtg tcacccacct ggcaatcaaa tctacagtca cgataaactt 720
tcattttttg aaatcgacgg ccgcaaaaac aaaagctatg ctcagaatct atgcctgctt 780
gccaaacttt ttctggatca caagactctt tactatgaca cggatccatt tttgttctat 840
gtgctaaccg aagaagacga gaagggtcat catatagttg gatacttttc aaaagaaaaa 900
gaatcagctg aagaatataa tgttgctgtg attcttgtgt tacctccatt tcaaaagaaa 960
ggatcacccg aaaaaccact atcagatttg ggacttctct catatcgatc gtactgggtca 1080
atggccatca tgaaagagct tttcgcattc aaaagacgac atccaggcga agatatcaca 1140
gttcaggaca tttcacaaag tacatcgatt aaacgagaag atgttgtgtc aacgttacag 1200
caacttgatc tatacaaata ctataaggga tcatacataa ttgtgattag tgatgaaaag 1260
cgtcaagttt atgagaaacg gattgaggct gcgaaaaaga agacacgaat taatccagca 1320
gctctgcaat ggcgacccaa agagtacgga aagaaaagag cgcagatcat gttctag 1377

```

<210> 17

<211> 458

<212> PRT

<213> Caenorhabditis elegans

<400> 17

```

Met Thr Glu Pro Lys Lys Glu Ile Ile Glu Asp Glu Asn His Gly Ile
 1      5      10      15
Ser Lys Lys Ile Pro Thr Asp Pro Arg Gln Tyr Glu Lys Val Thr Glu
 20      25      30
Gly Cys Arg Leu Leu Val Met Met Ala Ser Gln Glu Glu Glu Arg Trp
 35      40      45
Ala Glu Val Ile Ser Arg Cys Arg Ala Ala Asn Gly Ser Ile Lys Phe
 50      55      60
Tyr Val His Tyr Ile Asp Cys Asn Arg Arg Leu Asp Glu Trp Val Gln
 65      70      75      80
Ser Asp Arg Leu Asn Leu Ala Ser Cys Glu Leu Pro Lys Lys Gly Gly
 85      90      95
Lys Lys Gly Ala His Leu Arg Glu Glu Asn Arg Asp Ser Asn Glu Asn
 100     105     110
Glu Gly Lys Lys Ser Gly Arg Lys Arg Lys Ile Pro Leu Leu Pro Met
 115     120     125
Asp Asp Leu Lys Ala Glu Ser Val Asp Pro Leu Gln Ala Ile Ser Thr
 130     135     140
Met Thr Ser Gly Ser Thr Pro Ser Leu Arg Gly Ser Met Ser Met Val
 145     150     155     160
Gly His Ser Glu Asp Ala Met Thr Arg Ile Arg Asn Val Glu Cys Ile
 165     170     175
Glu Leu Gly Arg Ser Arg Ile Gln Pro Trp Tyr Phe Ala Pro Tyr Pro
 180     185     190
Gln Gln Leu Thr Ser Leu Asp Cys Ile Tyr Ile Cys Glu Phe Cys Leu
 195     200     205
Lys Tyr Leu Lys Ser Lys Thr Cys Leu Lys Arg His Met Glu Lys Cys
 210     215     220
Ala Met Cys His Pro Pro Gly Asn Gln Ile Tyr Ser His Asp Lys Leu
 225     230     235     240
Ser Phe Phe Glu Ile Asp Gly Arg Lys Asn Lys Ser Tyr Ala Gln Asn
 245     250     255
Leu Cys Leu Leu Ala Lys Leu Phe Leu Asp His Lys Thr Leu Tyr Tyr
 260     265     270
Asp Thr Asp Pro Phe Leu Phe Tyr Val Leu Thr Glu Glu Asp Glu Lys
 275     280     285
Gly His His Ile Val Gly Tyr Phe Ser Lys Glu Lys Glu Ser Ala Glu
 290     295     300
Glu Tyr Asn Val Ala Cys Ile Leu Val Leu Pro Pro Phe Gln Lys Lys
 305     310     315     320
Gly Tyr Gly Ser Leu Leu Ile Glu Phe Ser Tyr Glu Leu Ser Lys Ile
 325     330     335
Glu Gln Lys Thr Gly Ser Pro Glu Lys Pro Leu Ser Asp Leu Gly Leu
 340     345     350
Leu Ser Tyr Arg Ser Tyr Trp Ser Met Ala Ile Met Lys Glu Leu Phe
 355     360     365
Ala Phe Lys Arg Arg His Pro Gly Glu Asp Ile Thr Val Gln Asp Ile
 370     375     380
Ser Gln Ser Thr Ser Ile Lys Arg Glu Asp Val Val Ser Thr Leu Gln
 385     390     395     400
Gln Leu Asp Leu Tyr Lys Tyr Tyr Lys Gly Ser Tyr Ile Ile Val Ile
 405     410     415
Ser Asp Glu Lys Arg Gln Val Tyr Glu Lys Arg Ile Glu Ala Ala Lys
 420     425     430

```

Lys Lys Thr Arg Ile Asn Pro Ala Ala Leu Gln Trp Arg Pro Lys Glu
 435 440 445
 Tyr Gly Lys Lys Arg Ala Gln Ile Met Phe
 450 455

<210> 18
 <211> 9890
 <212> DNA
 <213> *Caenorhabditis elegans*

<400> 18
 tttcaaaaaa aaaaaattac ctcgtaatt tcaactctcct cgatgcgatg attatcctcg 60
 tccatttttac ctgaaaagtg tgattttttc acgaataaaa ttatttttcag atactttctag 120
 aaaaaaaaaa ctgaacggaa tgttacgaaa ttaatttttca aagttgcgaa actgaatttt 180
 cgacaaaaag tttcactgat attcattttca agcatattgc aacgttttta aattaatttc 240
 taagagaaaa aactgcaaaa caattcgaaa ataattttta caagttactt ttcgaaaaag 300
 taacaaaaat ccactaatga acaagaaatt tttgaacaaa aagagcttct caggctattt 360
 ttggacgaat attttaataa aactttaaaa aaatcaacga aaatccccta aaaatcgctg 420
 aaaattccaa aaattaaagt tcattctcga ccacacctct cgtaaatacag cactgagactc 480
 acgcaacgcg accgcgccgc actcaacggc attgagtaat gcggagcggc agcgtcgctg 540
 cgtctatttg tgtgtgtgtg cgattgtgtg tgggtgcgacg tggccgctct gtgtgcctct 600
 ctagtgagtg ttttccgacg agagacaaca cattttcgag agacgaagag agtggcgacg 660
 aggaagatag tgtggttaaga ggagagtgtg cgcgagggaa agagagcaaa gtgtgagtgt 720
 ctgtgagaag agaaggagac ccccccccc cgcgcgtcaa ccagtcgata gttggcctga 780
 gtgtaggggc ttctgttgta ttccactgct aacccccccc aaacacacaa aaagactcaa 840
 aaagtactgc ttaaaacaca gtgctcagct catttcattt ttgattttta tgctcgccgt 900
 catcggcgga tgaattcatc gcaaagtcgg tggcgattca acacgtgcgg cgtcctcgcc 960
 gctcttctta accgtagtta caacgtggga gtacagaaag atggccacta cttcgaaggc 1020
 gtttcgagcc cgggcgctcg actcgaaccg gtctatgact gtatactggg gccacgaact 1080
 tccggacctc tcagaatgca gtgttggaac cgggcgggtg acacaaatgc cgtctggcat 1140
 ggaaaaagaa gaagaacagg ttgggttttg gtggattatg gattactgct ccattttgaa 1200
 atttttcgag ttttaattgc ttttttcgaa ttcttggtgc ttttttctat ccgaatcatg 1260
 ttttaattcc gttttccgac tactttgaag aattttcaaa tttttgatcc ctgatgacgt 1320
 cactattttt gtctttgcct ttctggatcg cttttatagt tattttcatt ttttatttct 1380
 tttttacact tttaaactta acaattctct taattcatcc tattctattt aattttaagt 1440
 tttgattttt gatttttgat ttttctcttt tctcttttag ccgcgggtgg gcctttatta 1500
 caactcttaa atcataaaaa aaatcagttt aagcagttat acataactct tattatgaaa 1560
 aaatcgttat ttttcgacgg aaacttcata ctttgaattt atttccaatt tagattttat 1620
 tttctcaaag tcagctcaat taactaactt aaaatgtttt gtccatcccg caaatgttt 1680
 ttttttaata ttttaattct attttaattt ttggctttta aaaatcattt tgctaagcct 1740
 gagatgaagg cgaaatctcg agaaaaagca tttaaaaagt aataaattcc gttaaaaacg 1800
 actttttcta tcacagaaag tgttctctga gtgctaacaa ctttcttctg tccaaatttt 1860
 gacacaattt cccaattatg ccgacttatt acaccttttt ccgtcaatct tctagttttt 1920
 ccaccctct tgaccctgg tgacgtcatt tgtttgttct tcttccaaga catgccctgt 1980
 ggggtatttt ttctcaaaat ttttgcaaat ttattggatt ctaataaaaa ttccaggagt 2040
 ctagcaccag gaataataat gcaaatttga aaaaaaaatt aaacagaaat aatgatttta 2100
 aatgattatt taaattttta attttaaatt tccaggaaaa acacctgcaa gaagcgattg 2160
 ctgcccagca agccagtaca tcgggtattc agctgaacca tgtcattcca actccaaaag 2220
 tcgaccgagt cgaagatcaa cgctatcact ccacttatca caacaagaat aaaatgcacc 2280
 gttcaaagta tatcaaagtt catggtgagt ttttttaacc aaaatttcgg cgaaaataat 2340
 ttaatttcgg gttttttgaa attaatctcc gcttgggttt tcttgtattt attatttttt 2400
 caaattcctc tctgaattcg aaagaaaata acttgatttt tcagacttcc tggctaaaaac 2460
 cttcaaaaat gtttgttgat tgggtccaaa ttttcgcctg attccgaatt tcgatgtgac 2520
 aaattcaaaa aaaaattccc tgattttata ttcaagcttt gtgtttgtgt gttctttttg 2580
 gagcgcgctt gcatcgtttg attttcttcg tcttttttaa aatttatttt cgcttgtttc 2640
 attcattttt gtcgagtttt tttctgccaa aatgaatgaa actggtttta aaaattgaat 2700
 tcggcgaaaa taaattttga aaaacgaaac aaatcaaacg atgcaagcgc gctccaatgc 2760

gattttttt	ggcgcgga	ttcgtgatt	caagcttaa	tataaaatca	ggtatatttt	2820
ttcgacttt	ttcacgttg	aattcgga	cagaggaaa	ttttgagtca	atcaaaaata	2880
tttccagat	ttcggatat	ttaatgcac	aaaaatga	tttcaccccc	atactcccag	2940
aaaaataag	aaacaaatt	cgaaatat	ttccctgat	aaatttttt	tttttttaac	3000
tacacttct	tgttttga	tgagaaagt	catttttct	cgtttcttat	cagttatcat	3060
ttgaaaagg	tcagaattt	atgacgat	atttgtttag	ttacctccct	tttttctgaa	3120
cagtttttg	gaaaaaagg	gaaaaaccg	aatttttct	gaaaaatgt	tttattttca	3180
gcttggaag	cactcgaac	agacgaacc	gagtatgac	acgacacaga	agatgaagca	3240
tggctatcag	atcacactc	cattgacccg	cgcgttttg	aaaagatat	cgacacagtg	3300
gagagccatt	catcggagac	acagatcgc	agcgaagatt	cggtgattaa	tttgcataaa	3360
tgtaagttg	cgaaatttc	attgaaacc	cccccccca	aaaatatcgt	ttaattgcag	3420
cactggactc	atcaatcgt	tacgaaatat	acgaatat	gctgtcgaag	cgaacatcgg	3480
ctgcgacgac	gtctggttg	gttgagtg	gtggattaat	tccgagagtc	aggacagaat	3540
gtcggaagg	aagaatttg	ctattttga	cgaatttcgt	gatgaaactt	ctctaaaact	3600
tttaaagtt	tttatggcg	ttcaaaatt	cggaaaatt	acactgattt	tagctaaaaa	3660
cttgaattt	ggtcatttg	ccgtgtcaca	tctgtccgaa	atcgactttt	tttggaaatta	3720
tcactcctta	ttgcacatt	ggctagttaa	tctcatttaa	tttcgttgat	tactaaggta	3780
cattttaaac	caataggtaa	ccaacaaaa	actatcataa	tttttctaca	ctttttaatt	3840
ttccgacact	acttgaata	cccataagt	gaccaatttt	gatagttttt	ggctggttac	3900
cggcttttaa	tgtaccttat	taatcaacaa	aattaaatga	gataaactag	ccaaatgtgc	3960
aataaaggat	gataattcca	taaaaagtc	attttggaca	gatgtgacac	gggcaaatga	4020
ccaaaattca	agtttttagc	taaaatcagt	gtatttgttt	cgaagttttg	aaccgctata	4080
aaaaaatttt	tggaaatgct	ttggcaagtt	tcattacgaa	attcactcat	tttctatacg	4140
caaaaattag	aattttcaat	taaaaattca	ttttacagga	tggacaaggt	gttatcaatc	4200
cgtagcttg	attccgtcga	cgtgccgaga	aaatgcagac	tcgaaagaat	cggaaaaaac	4260
atgaagatte	gtatgagaag	attctcaagt	tggtagatga	catgtcga	gctcaacagc	4320
tcttcgatat	gactgcccg	cgagaaaagc	agaagctcgc	gttgattgat	atggaatcgg	4380
agatttttag	gaaacgaatg	gagatgtcag	attttgggtg	ttctccgagt	tcgttcaatg	4440
agatcaccga	aaagattcga	gcagcagcaa	cgttgggaag	cgtgaaacca	ccactggcag	4500
aaatcaacgg	atcagatgaa	gtgaagaaga	ggaagaagcc	gagacgaaag	attgctgata	4560
aggattta	atcgaaagcc	tggcttaaaa	agaatgcaga	aagttggaat	cggccgccgt	4620
cgctcttttg	acaacacagt	ggaaatgttc	cgacgggtac	aacgaagcca	gttcgagagt	4680
cgttggcgaa	tgggcgattt	gcgttcaagc	ggaggagagg	atgtgtttat	cgcgcggctc	4740
tcaccgttta	caatgtgcct	acagcgcttg	ctacagtacc	tccagtacag	actcaagcag	4800
cagtggcttc	atcatcatcg	tcaaaatcaa	cggatatggt	gccgtcgaac	atgaagtctt	4860
ttgaaacttt	tgttcgggat	tcacaggatt	cagtttctcg	atctcttggc	tttgtacgcc	4920
gacgaatggg	acgaggtggg	cgagtgtgat	tcgatcggat	gcctcgcaat	cgagacgaca	4980
acgacgaacg	cacttcgaca	gatccatggg	ccgagtattg	tgctcgcgat	agttcaaggt	5040
gagatttttg	aataagaatc	ttaatttcac	gagatttttg	tttttttcgc	tgctttttct	5100
gtaattttgt	ggtatttttt	ctcgtatttt	caattaaaaa	acgggtttta	aataattttta	5160
acctgaaatt	tcgctaaaaa	ccaagaaatt	tcattaaaaa	atgcaacaaa	aaaaaagact	5220
ggaggcacca	ccgaatggag	aacaggagaa	cccaaaacca	cgccattttt	tccgtgccgg	5280
gcggcgaaaa	tttttgcgag	atttgctgca	atttttcggt	ttacaaacga	aacaacgaag	5340
ctctgaatgt	gttattttcg	agcttcggtg	tttcgtttgt	aaaacgaaaa	attgcagcaa	5400
tttctgcaaa	aatttgcgcg	cggcacggaa	aaatgggcgt	agtttttaggt	tctcctgttc	5460
tccttttggg	ggtgcctcca	gtctttttcg	cattctta	gaaatttctt	tgtttttttag	5520
cgaaatttca	ggttaaaatt	atttaaaacc	cgtttttttt	tcaattggaa	atgcgaggaa	5580
aaaccacaaa	atcacagaga	aagcttttgg	attttttcgc	agctttttct	gtgattttgt	5640
ggtttttcct	cgcattttca	attgaaaaaa	aaacgggttt	ttaataattt	tcacctgaaa	5700
tttcgctaaa	aacgaggaaa	tttcattaca	aatgcaaaaa	agactggagg	caccaccgaa	5760
accgaatgca	gctcagaaca	ggatttacca	aaacaggatg	cagtaggcgg	agccaattcg	5820
caaccaccgc	atgcttat	cgcatgcctc	gcacgttttt	tttttctctt	gaaacaatgc	5880
aacaatatca	aggaaaaaac	gtgcgagact	tgcgaaataa	gcatgcggtg	gttgcgaatt	5940
ggctccgccc	actgcattct	gttttggtaa	attctgttct	gagctgcatt	ctgtttttgt	6000
ggggcttcca	gtcttttttg	tgcattttta	atggaatttc	ttcgttttta	gcgaaatttc	6060
aggttaaaat	tatttaaaac	ccgttttttt	ttcaatttga	aatgcgagga	aaaaccacaa	6120
aatcacagag	atagcgaggc	cccacgaaaa	ggggagcaga	acaaaaaagg	gggggggggg	6180
gctggcactg	tgccaaacgc	acaaaacgct	ttttattctt	attcaacgca	cgactttgtt	6240

ataaccacac	tccgttatta	cgcacgcgc	gctgttttagc	gtgaaaatac	aaaaaaacgt	6300
cgtgcgttga	atgagaataa	aaaagcgttt	tgtgcgtttg	gcacagtgcc	agctctcctt	6360
ttcgcagatc	cccttttcgt	ggggcctcag	agaaagctgc	cataaaacttt	tttcttcgcg	6420
ctaagaccaa	taccaataaa	tccttgcgcc	tttaatatgc	aaactatatt	tttcttccag	6480
aaccttccgt	gtcgaaca	gttcgcttgg	taccgaagaa	gaaaccgatg	atctaagccc	6540
gaaatctctg	tatttcgctc	gcagtaatcg	gttcgcattc	aacgatgatg	aaactgaacg	6600
ggaatggact	tcaagatgcc	aacaatcatc	gtggagagat	acagagggtg	atgatgagct	6660
gaaaaagcgg	gaaacaacgt	ctgaaagtga	gattttgaac	gattttacctg	ggaaaataga	6720
ttattttggg	cctattttta	ttatttaatt	gcagaattta	ccgaaaccac	gacgaatgga	6780
agtaccaaaa	cacacacaga	atcggatgat	agtgaagtgtg	aacggatgga	ggttgatgat	6840
caagttgatg	aagctcaa	aactgtatca	tcataaaaag	acgatggaat	gaatggaaat	6900
gataagaacg	aggatgaaga	agatgatgat	gatgatattg	atgtagatga	acatcagact	6960
gtcgtgggtg	tgcacagca	ccagcagcag	cagcatcacc	agcaaaaagt	tcggcatcaa	7020
atgaatggtg	gtggtggtg	tggtggagtg	gtaaaactga	aaccgccgct	gcaagaactt	7080
tcgccgccgc	tttcgggaaa	cggaagagcg	gacagagcgg	aaccgacgcc	ggttccggca	7140
aaggtagtga	ggcttttttt	ttaaatactc	gaaaaagaag	gaaaaaatcc	cactttttaa	7200
aatacgattc	ttaaaaatgc	gaattccctc	caaaatgaga	actctgattg	gccagggagc	7260
tctcattttg	aatggaaatt	agtcaaaatt	gaaaaatccc	gttttttttt	taagttggat	7320
ttttcatttt	ctcgcgattt	tttcgcggtt	tctgtgcat	tcctgaattt	aaactttaat	7380
aaattaaaaa	tgtctggaat	attgacaaat	tatgtctcaa	atttttttgcg	cgggagttca	7440
aaaataattt	ggcccttttt	attttttatt	ttgcaaaaat	atataaaaaa	tcatttttaa	7500
aaatttagaa	acatttttta	atttttttaa	cagttatatt	cgctatattg	ggacggattt	7560
ctgtcattaa	acttgggtgt	gtcgaatttt	ttttattgct	ttataagact	caaaattgtc	7620
tgaaaacacc	gaattttata	atgaaacttc	ttggaaactt	ctcaaaaaaa	agttatgacg	7680
gctcaaaaaa	tgacctaaaa	tttgtaaata	tttgaaattt	gacttgtcgc	aacggctgga	7740
aacaattttt	ttttttgaaa	tcaccgtcaa	atttttgagta	taaaatttaa	ttatttttgcg	7800
ttttcaactc	gatttttggt	attttcaagt	cgatggacgg	caagatttgg	ttaaaaaatt	7860
aaaagccgtc	cattttctcg	ccgtccattg	actttaaact	acctaatacg	agttgaaaac	7920
gcaagataat	tgacatttat	accctaaatt	tgactgtggt	tttaaaaaag	ttagtttcca	7980
gccgctgcga	caagtcaa	ttccaatttt	aactatttta	ggccattttt	tgagccatca	8040
taactttttt	ttgagaagtt	tttaagaagt	ttcatcatga	aattcgggtg	tttcagacaa	8100
ttttgagtct	aataaagtaa	ttttaaaaaa	ttcgacagac	accaccttta	tagcaatttt	8160
gaattttttt	ttaaacttgt	cttgaaaaat	cttgaaaaaa	gtcgaataaa	ttcccathtt	8220
cctattttct	tttttgca	tgtgcggaac	ggtgtcggac	tcagatgatt	ggagagagcc	8280
gagtggatca	ccatcagaat	cgaattcatc	aaccgaatgg	ggtggctata	cgccacaaga	8340
acagcatgca	gttggtgttg	ccaacgcgg	agctgtcgct	ttcaaggaaa	aattgatgaa	8400
tggcgtggat	gatgatgatg	atcaacaacc	atcgcgggct	agaggagcac	gagatcattc	8460
catcaagag	ttcgttagtt	tttctttgct	tttttttttt	ttgatttttg	agagcaaat	8520
tgaaaagt	tacacggttt	ttgaaaaact	gttgaaatta	aaatttgttg	agaatttgat	8580
ttcgagcaag	ttttattttt	aaaaaattga	atttttcaga	aaattctgag	ttttcttttt	8640
aaaaaattga	aattttcaga	aaattctgag	tagcaagaat	ctttaagatc	cttaatttct	8700
atgcaagaat	acgtaggagt	tttactttgc	tcaggaaatt	ttattttttg	tcagaggagt	8760
atatccgaaa	aagaacaaaa	aaaatgcaca	tttctcaaaa	cgcgtatttt	tttttcagtt	8820
cgatgtcaac	ggtaacactg	ctggaacgga	aaaagttcat	gatgccgtcg	acaatcggtc	8880
tataatttga	actctctgct	gctgcttctg	ctactgctgc	tactgctgct	catcgccaat	8940
tttcaatcct	cctgagattt	tttgatggtc	attcattggt	ttgtgcatat	ctctctctct	9000
ctctctctct	cccatgattc	tcaaatattt	caatgtattt	acacccccac	tctgtccgct	9060
gcctaattccc	cgaccgaata	atcagattcg	ctggaaaaat	ctgcgattct	ttaatattgc	9120
aaccacccac	ccaataatat	gtgtctcatc	atctcgggtac	tctcacttga	gccgtgtttt	9180
ctgtagtatt	ttattctcta	aaaaaaaaatc	attttttaata	taataatacgt	acacatttat	9240
atctgtaata	tatattttta	aaaatgattc	ccccctcccc	tccattcggt	gttttttttc	9300
tgtgggtttc	aagcttttga	gctgtgaaaa	atctcatccc	atcatcattt	tctattgttt	9360
tttttcacag	ttgaaatata	ctattttatc	tttttccctt	ttttttcatt	tttttttttt	9420
catcgtgcgg	gattcatttt	tcgtcccgcg	aaacgcccgc	cgccgcccac	tcccactctc	9480
tctctcagtc	tcttcttaac	gatcttcgaa	actattttta	tttccctcat	taacaattac	9540
gaggtcgtct	tttttttttc	ccacccccca	ctgttttggtg	taatttttgt	gttcggggag	9600
gttttttgtg	tgtggatttt	gggatttttt	ggaatttttc	aacaaaaaat	tccccgaaa	9660
tcaaaatttt	ttcccathtt	ccctcaata	ttagtactgt	tgtataaata	aacttgctct	9720

```

ctctctctct ctcgaaatct cctactatta tttttttaaa agatttttcc aacaaaaatt 9780
caaaaaacca cacaaacgac ctctctgcac gcggtaatcc tctctctttt tgtcccccac 9840
tttctctgtt tctctttttt tctatcccct atacctgtga ttggaatata 9890

```

<210> 19

<211> 2388

<212> DNA

<213> *Caenorhabditis elegans*

<400> 19

```

atggccacta cttcgaaggc gtttcgagcc cgggcgctcg actcgaaccg gtctatgact 60
gtatactggg gccacgaact tccggaccta tcagaatgca gtgttggaac cggggcggtg 120
acacaaatgc cgtctggcat ggaaaaagaa gaagaacagg aaaaacacct gcaagaagcg 180
attgctgccc agcaagccag tacatcgggt attcagctga accatgtcat tccaactcca 240
aaagtcgacc gagtcgaaga tcaacgctat cactccactt atcacaacaa gaataaaatg 300
caccgttcaa agtatatcaa agttcatgcc tggcaagcac tcgaacgaga cgaacccgag 360
tatgactacg acacagaaga tgaagcatgg ctatcagatc acactcacat tgaccgcgcg 420
gttttggaac agatattcga cacagtggag agccattcat cggagacaca gatcgcgagc 480
gaagattcgg tgattaattt gcataaatca ctggactcat caatcgtgta cgaaatatac 540
gaatattggc tgtcgaagcg aacatcgggt cgcacgacgt ctggtttgtg tggagtcggg 600
ggattaattc cgagagtcag gacagaatgt cgggaaggat gacaagggtg tatcaatccg 660
tacgttgcat tccgtcgacg tgccgagaaa atgcagactc gaaagaatcg gaaaaacgat 720
gaagattcgt atgagaagat tctcaagttg gtacatgaca tgtcgaagcg tcaacagctc 780
ttcgatatga ttgcccagcg agaaaagcag aagctcgcgt tgattgatat ggaatcggag 840
attttagcga aacgaatgga gatgtcagat tttggtggtt ctccgagttc gttcaatgag 900
atcaccgaaa agattcgagc agcagcaacg ttggaagtcg tgaaaccacc actggcagaa 960
atcaacggat cagatgaagt gaagaagagg aagaagccga gacgaaagat tgetgataag 1020
gatttaatat cgaaagcctg gcttaaaaag aatgcagaaa gttggaatcg gccgccgtcg 1080
ctctttggac aacacagtggt aaatgttccg acggttacaa cgaagccagt tcgagagtcg 1140
ttggcgaatg ggcgatttgc gttcaagcgg aggagaggat gtgtttatcg cgcggctctc 1200
accgtttaca atgtgectac agcgccgtgt acagtacctc cagtacagac tcaagcagca 1260
gtggcttcat catcatcgtc aaaatcaacg gatattggtg cgtcgaacat gaagttcttt 1320
gaaacttttg ttccgggattc acaggattca gtttctcgat ctcttggtt tgtacgccga 1380
cgaatgggac gaggtgggag agttgtattc gatcggatgc ctcgcaatcg agacgacaac 1440
gacgaacgca cttcgacaga tccatgggac gagtattgtg tcgcgatag ttcaagaacc 1500
ttccgtgctc gaaacagttc gcttggtacc gaagaagaaa ccgatgatct aagcccgaac 1560
tctctgtatt tcgctcgag taatcggttc gcattcaacg atgatgaaac tgaacgggaa 1620
tggaactcaa gatgccaaac atcatcgtgg agagatacag aggtggatga tgagctgaaa 1680
aagcgggaaa caacgtctga aaaatttacc gaaaccacga cgaatggaag taccaaaaca 1740
cacacagaat cggatgatag tgaagttgaa cggatggagg ttgatgatca agttgatgaa 1800
gtcaaataa ctgtatcatc atcaaaagac gatggaatga atggaaatga taagaacgag 1860
gatgaagaag atgatgatga tgatatggat gtagatgaac atcagactgt cgtgggtgtg 1920
catcagcacc agcagcagca gcatcaccag caaaaagttc ggcatacaat gaatggtggt 1980
ggtggtggtg gtggagtggt aaaactgaaa ccgccgctgc aagaactttc gccgccgctt 2040
tcgggaaacg gaagagcgga cagagcgga ccgacgccgg ttccggcaaa gatgtgcgga 2100
acgggtgctg actcagatga ttggagagag ccgagtggat caccatcaga atcgaattca 2160
tcaaccgaat ggggtggcta tacgccacaa gaacagcatg cagttgttgt tgccaacgcg 2220
gtagctgtcg ctttcaagga aaaattgatg aatggcgtgg atgatgatga tgatcaacaa 2280
ccatcgccgg ctagaggagc acgagatcat tccatcaaag attcgatgtc aacggtaaca 2340
ctgctggaac ggaaaaagtt catgatgccg tcgacaatcg gtctataa 2388

```

<210> 20

<211> 795

<212> PRT

<213> *Caenorhabditis elegans*

```

<400> 20
Met Ala Thr Thr Ser Lys Ala Phe Arg Ala Arg Ala Leu Asp Ser Asn
 1          5          10          15
Arg Ser Met Thr Val Tyr Trp Gly His Glu Leu Pro Asp Leu Ser Glu
          20          25          30
Cys Ser Val Gly Asn Arg Ala Val Thr Gln Met Pro Ser Gly Met Glu
          35          40          45
Lys Glu Glu Glu Gln Glu Lys His Leu Gln Glu Ala Ile Ala Ala Gln
          50          55          60
Gln Ala Ser Thr Ser Gly Ile Gln Leu Asn His Val Ile Pro Thr Pro
65          70          75          80
Lys Val Asp Arg Val Glu Asp Gln Arg Tyr His Ser Thr Tyr His Asn
          85          90          95
Lys Asn Lys Met His Arg Ser Lys Tyr Ile Lys Val His Ala Trp Gln
          100          105          110
Ala Leu Glu Arg Asp Glu Pro Glu Tyr Asp Tyr Asp Thr Glu Asp Glu
          115          120          125
Ala Trp Leu Ser Asp His Thr His Ile Asp Pro Arg Val Leu Glu Lys
          130          135          140
Ile Phe Asp Thr Val Glu Ser His Ser Ser Glu Thr Gln Ile Ala Ser
145          150          155          160
Glu Asp Ser Val Ile Asn Leu His Lys Ser Leu Asp Ser Ser Ile Val
          165          170          175
Tyr Glu Ile Tyr Glu Tyr Trp Leu Ser Lys Arg Thr Ser Ala Ala Thr
          180          185          190
Thr Ser Gly Cys Val Gly Val Gly Leu Ile Pro Arg Val Arg Thr
          195          200          205
Glu Cys Arg Lys Asp Gly Gln Gly Val Ile Asn Pro Tyr Val Ala Phe
          210          215          220
Arg Arg Arg Ala Glu Lys Met Gln Thr Arg Lys Asn Arg Lys Asn Asp
225          230          235          240
Glu Asp Ser Tyr Glu Lys Ile Leu Lys Leu Val His Asp Met Ser Lys
          245          250          255
Ala Gln Gln Leu Phe Asp Met Thr Ala Arg Arg Glu Lys Gln Lys Leu
          260          265          270
Ala Leu Ile Asp Met Glu Ser Glu Ile Leu Ala Lys Arg Met Glu Met
          275          280          285
Ser Asp Phe Gly Gly Ser Pro Ser Ser Phe Asn Glu Ile Thr Glu Lys
290          295          300
Ile Arg Ala Ala Ala Thr Leu Glu Val Val Lys Pro Pro Leu Ala Glu
305          310          315          320
Ile Asn Gly Ser Asp Glu Val Lys Lys Arg Lys Lys Pro Arg Arg Lys
          325          330          335
Ile Ala Asp Lys Asp Leu Ile Ser Lys Ala Trp Leu Lys Lys Asn Ala
          340          345          350
Glu Ser Trp Asn Arg Pro Pro Ser Leu Phe Gly Gln His Ser Gly Asn
          355          360          365
Val Pro Thr Val Thr Thr Lys Pro Val Arg Glu Ser Leu Ala Asn Gly
          370          375          380
Arg Phe Ala Phe Lys Arg Arg Arg Gly Cys Val Tyr Arg Ala Ala Leu
385          390          395          400
Thr Val Tyr Asn Val Pro Thr Ala Pro Ala Thr Val Pro Pro Val Gln
          405          410          415
Thr Gln Ala Ala Val Ala Ser Ser Ser Ser Lys Ser Thr Asp Met
          420          425          430
Val Pro Ser Asn Met Lys Phe Phe Glu Thr Phe Val Arg Asp Ser Gln
          435          440          445
Asp Ser Val Ser Arg Ser Leu Gly Phe Val Arg Arg Arg Met Gly Arg

```

450		455		460
Gly Gly Arg Val Val	Phe Asp Arg Met Pro Arg	Asn Arg Asp Asp Asn		
465	470	475	480	
Asp Glu Arg Thr Ser	Thr Asp Pro Trp Ala Glu Tyr Cys Val Ala Asp			
	485	490	495	
Ser Ser Arg Thr Phe	Arg Ala Arg Asn Ser Ser Leu Gly Thr Glu Glu			
	500	505	510	
Glu Thr Asp Asp Leu Ser	Pro Lys Ser Leu Tyr Phe Ala Arg Ser Asn			
	515	520	525	
Arg Phe Ala Phe Asn	Asp Asp Glu Thr Glu Arg Glu Trp Thr Ser Arg			
	530	535	540	
Cys Gln Gln Ser Ser	Trp Arg Asp Thr Glu Val Asp Asp Glu Leu Lys			
545	550	555	560	
Lys Arg Glu Thr Thr	Ser Glu Lys Phe Thr Glu Thr Thr Thr Asn Gly			
	565	570	575	
Ser Thr Lys Thr His	Thr Glu Ser Asp Asp Ser Glu Val Glu Arg Met			
	580	585	590	
Glu Val Asp Asp Gln	Val Asp Glu Ala Gln Ile Thr Val Ser Ser Ser			
	595	600	605	
Lys Asp Asp Gly Met	Asn Gly Asn Asp Lys Asn Glu Asp Glu Glu Asp			
	610	615	620	
Asp Asp Asp Asp Met	Asp Val Asp Glu His Gln Thr Val Val Gly Val			
625	630	635	640	
His Gln His Gln Gln	Gln Gln His His Gln Gln Lys Val Arg His Gln			
	645	650	655	
Met Asn Gly Gly Gly	Gly Gly Gly Gly Val Val Lys Leu Lys Pro Pro			
	660	665	670	
Leu Gln Glu Leu Ser	Pro Pro Leu Ser Gly Asn Gly Arg Ala Asp Arg			
	675	680	685	
Ala Glu Pro Thr Pro	Val Pro Ala Lys Met Cys Gly Thr Val Ser Asp			
	690	695	700	
Ser Asp Asp Trp Arg	Glu Pro Ser Gly Ser Pro Ser Glu Ser Asn Ser			
705	710	715	720	
Ser Thr Glu Trp Gly	Gly Tyr Thr Pro Gln Glu Gln His Ala Val Val			
	725	730	735	
Val Ala Asn Ala Val	Ala Val Ala Phe Lys Glu Lys Leu Met Asn Gly			
	740	745	750	
Val Asp Asp Asp Asp	Gln Gln Pro Ser Pro Ala Arg Gly Ala Arg			
	755	760	765	
Asp His Ser Ile Lys	Asp Ser Met Ser Thr Val Thr Leu Leu Glu Arg			
	770	775	780	
Lys Lys Phe Met Met	Pro Ser Thr Ile Gly Leu			
785	790	795		

<210> 21

<211> 37007

<212> DNA

<213> *Caenorhabditis elegans*

<400> 21

```

cagctgatgt tgttgatgga aaaatgacgg ctgcaaagaa gccattggct gcaactgagc 60
caaaagtgca taataaataa atgtgtttct aggatcttct aataattttt tttctgtttt 120
ctagctctaa acttgatattt atttcattct tgttctacca aattcccacg gattctacgc 180
tttatgtttc taaattatta ttctttttta tttatatctg cattttcttc taaaaactct 240
ggtcattttc ttgttttttt ctgggtaatt ataaaaatta gtcatacaaa tcttggttaa 300
tatctggcta ttcagtgaac aaaccatttt ccgctctaaa ttcgaccoga atcaatcgaa 360
aaatggctca aaacgatgcc atctggctgc aacccccctg tcgtctctca attttgtgta 420

```

```

ctctctcgca gccacgcacg cgacgcacag cactcgcgtc gcggtcgcag ttcttttttca 480
aatttatcgc gccatttttg ttttgccctca tatttatcgg ctcacgattg attttcgtcg 540
aaaaacgcgc ttaatcgatt cctttttacc tgaaaaatgt tgttccaatt ggaaaaccag 600
ttgaagatcg atgaattttc aagaaaatca ttcaaatagg caaaacccgc tgaactttga 660
aattcgattt ttgagttttt tgaagaaaat ataattattt catcatttat gttggctctg 720
ttggctctca gcatagaaaa ttccgacatg acattagaaa ttcataataa ctgctcccaa 780
tatcgggatt agaacgattt tcagctcaaa atatggaaaa ttggttacat aaaccgcata 840
ttttagcat taatcttgaa cagctatatg gcattaaaaa aaaatatata tatacattgt 900
ttttctctc gaagttttctc tttttgtttc taaaatccgg aatataattt aaaaaaccac 960
ataaatttca atttgcagta cgagttcccc ccgaatcaca atgcccggca caccgggtgc 1020
tgcttcaagt actcgaataa gcagacgtac atcatcaaga tcagtggctg atgatcagcc 1080
atcgaattcg tctcgggtgg cccacctoc ttcacccatt gccatagaaa ctgatgaaga 1140
tgcggtagtt gaggaggaga aaaagaagaa aaagacatca gatgatttgg aaattatcac 1200
tccaagaact ccagtcgata ggcgaattcc ctacatttgc tcgattcttt tgactgaaaa 1260
tcgatcgatt cgcgataaat tgtacgattt tttaaattta attactttcc tcaaatccga 1320
ataattatta gatcgcgctt cgcgtttctg catccgcggt attttgcctt cccactgaaa 1380
atagcagatt tatcgaattt ttagcttaaa aaaaaaatgt tttttctgca tttttcaaac 1440
aaaccttttg taaaacagtg aaaatcgaat ttcaaatagc taaaatgaat ttttttttg 1500
tccactgggt gtggaatggg ttgaatttga agaaatcagc gggatttttc gtattttctg 1560
aatatttttc tattaataat cggtttcaaa cacttttttg acttttgaat agaaaaatat 1620
tgagaaaata cgaaaaatcc agctaacttc cagcttggtc aaattcaaac cattccacaa 1680
ccagtggacg aaaaaagttc attttagtca tttgaaattc gatttggttt gtttgaaaaa 1740
tgcaaaaaaa aaatatTTTT taaagctaaa aatttgataa atctgaaaaa aatctgctat 1800
tttcagtggg aaggcaaaat accgcgaagc gcagcaagcg cgtcttaata attattccgc 1860
ttcgagaaga gcgtgtatta tttcattgtt acatttcaaa attatgaatt aatgtttttc 1920
agggttctga gcagcgggtcc agttcgtcaa gaagatcacg aagaacagat tgctcgagct 1980
caacggatac agccagttgt cgatcaaat caacgagtcg agcaaatgta tgtgaagctg 2040
aaaaattgca ccacaaatca attattctaa tcttgtttta cagcatactc aatggttcag 2100
tggaagatat tctgaaagat cctcgattcg cagtaatggc agatctcaca aaagaaccac 2160
caccaacacc tgcacctcct cctccaatcc agaagacaat gcaaccgatt gaggtgaaaa 2220
ttgaggatt agagggtca aatacggctc aaccgagtgt tctgcccagt tgtggaggag 2280
gagagacgaa tgtggaaaaga gccgccaaaa gagtgaattt tgaagataga ttggtgtgta 2340
aaaaatgaat gtttatatat tcaactgcaac ttttctctca cgagggacga ggaaaagtgg 2400
tttctaggcc atggccgagg tgccgacaag tttcagcggc catttatctt gctttgtttt 2460
ccgcctgttt tctttcgttt ttcacgatt ttttctgttt tttcttaata aaactgataa 2520
ataaataattt tttgcagatg ctaaaacaat ttccaagtaa aaaaattatg tattcagtg 2580
gcaagcagcg gtgaaagtgg tcaatgcaat atgatggatt acgggaatac aaaacctaaa 2640
ctttttctga aacatgatac atacgtgct taaatgctga gactacctga ttttcataac 2700
gagaccgctg aaaaagtttt gaggttttca aaattcaaat tttttggtga aaaagtcgag 2760
attttcgcac aaaaagttag attctgaaaa cctcaaattt ttttcagcgg tctcgttatg 2820
aaaatcagg aatttcagca tcatatgtat catgtttcaa aaaaagttaa ggttttgtat 2880
tccgtaatc catcatattg cattgaccac tttcacgcgt gcttgcccac tgaatacatg 2940
attttttact tggaaattgt tttagcatct gcaaaaaata tttatttatc agttttatta 3000
agaaaaaacg aaaaaaatcg gtgaaaaacg aaagaaaaca ggcggaatac aaagcaagat 3060
aatggccgc tgaaacttgt cggccctcgc gccatggcct agaaaccact tttcctcgtc 3120
cctcgtgagg aaaaagtgtc agtgttattg taaatctcac aagagtctgg catgatttct 3180
caaaggcgca tggatttatt cagccctaaa attaaataaa tccatacgac tttaaagggtg 3240
gagttcggaa aatgaggatt ttacttttaa atgctcaaac tagtcccaa tgccgaatta 3300
ccacaaaaga aaacggaaa aaaattcatc aagtttgaaa aaaatgcgga tgattttgtt 3360
gaaatttcaa cgctcgctaa tttcctaatt ttgaaccgcg cttttgtccg cgccgcactc 3420
gtagaattg catccgcgct gtttccctcc tcttcggcg ccctacttct tttcgattgg 3480
aaatgatgaa aaatgagac aaaactagaa ttcacgtagc gcgtcggaat tgatgaaaat 3540
atcatggatg cagcagatct acggagtgcg gcgcggacaa acggcgcggt aattcaaatg 3600
aggaatatta gcgagagttg aaatttcaac aaaatcagcc gcattttttt caaacttaat 3660
gtattttttt tcgtttttct tttgtagtaa ttcggcattt ggggctagtg taagcatttt 3720
aaagtaaaat cctcattttc cgaactccac ctttaaaggt ggagtaccga aatttgagac 3780
tttgcttttt taggcccaaa ttggtccaaa actaccgaat tttgtaatga gacgttctga 3840
aaatttatcc aaaaaatgtt atggcggttc aaagtccggc aaaatagggc ccattttcag 3900

```


ctaaaatcaa	atTTTTTTTT	ccaactTTTT	cggtgtcgca	acgtctggag	cctaattttt	3960
atttattaat	cactTTTTta	taaatattgt	agcctttgat	taggcgttta	ttcgctgatt	4020
taagtacatt	tatggTTTT	ggggcacaaa	taaaagtttc	atTTtatgcc	ccaaaaacca	4080
taaatgtact	taaatcagcg	aataaacgcc	taatcaaagg	ctacaatatt	tattaaagag	4140
tgatgaataa	ataaaaaatta	ggttccagac	gttgcgacac	cgaaaaagtt	ggaaaaaatt	4200
ttgatttttag	ctgaaaatgt	gccttatttt	gccgcgaact	ttgaaccgcc	ataacttttt	4260
ttgagaaaga	aattttcaga	acgtctcatt	acgaaattcg	gtagtTTTaa	accaatttgg	4320
gtctaaaaag	tttcaaattc	caataaaaaca	taccaaaagtc	ttgtgaaatt	acaataaaact	4380
attcctaatac	gtattataat	ccattctcaa	ttcttgcagg	aagcgcatgt	attggctcga	4440
atcgccgagc	tccgtaagaa	cggcttatgg	tcgaacagtc	gtctgccaaa	gtgcgtcgaa	4500
cctgaacgta	ataaaacgca	ttgggattat	ctactggaag	aggtcaaattg	gatggcagtt	4560
gattttccgaa	ccgagacgaa	tacgaagcga	aaaatcgcca	aagttatagc	tcacgccatt	4620
gcgaaacagc	accgcgacaa	gcagatcgag	attgagagag	ccgccgaacg	ggagatcaag	4680
gagaagcgaa	aaatgtgtgc	aggaatcgcg	aagatggtac	gggattttctg	gtcgtctacg	4740
gataaagttg	tggatattcg	agcgaaggaa	gttctggagt	cgaggctcag	gaaggcgaga	4800
aataagcatt	tgatgtttgt	aattggacaa	gtcgatgaaa	tgagcaatat	tgtagcaaga	4860
ggacttgttt	catcgtcgaa	atccccatca	attgcatcgg	atcgagatga	taaagatgaa	4920
gaattcaaag	cacctggctc	tgattcagaa	tctgacgatg	agcagacaat	tgcaaacgcg	4980
gaaaaagcac	agaaaaagga	agatgttcga	caggaaagttg	atgctcttca	aaacgaggca	5040
actgtggata	tggatgactt	tttgtacact	ttaccgccgg	aatatctgaa	ggcttatggg	5100
ctgacgcagg	aggatttgga	ggagatgaag	cgcgagaaat	tggaggagca	gaaggctcgg	5160
aagggaagctt	gtggtgataa	tgaggagaaa	atggagattg	atgaagttcg	taggatgctc	5220
ctaaaaaaat	tacctaaaaa	aaatcgattt	tccctggaaa	aaatcctctg	gaaatgaccc	5280
gaaacgcat	ggcggctcga	aattttgaaa	aaaaaaaccc	cccaaatttc	cagctaaaaat	5340
ctcaaatTTT	attgcatatt	ttggtagttc	ttttgttgtc	cgagggtcgt	ttttcagctg	5400
aaaatgtacc	tgaatctgca	agtaaacgac	caatatatgc	aataaatgat	gataattaat	5460
ttccgatact	gaaatgtggg	cgaattttga	gatttcgact	gaaaaacgtc	taaaaatcac	5520
ccaaaacccg	gctttaccgc	acgaaggttg	gaagaaaatg	gccaatTTTT	agccaaaatc	5580
tcaaatTTTc	tccacttttc	agtcagaaat	tagttttttg	aaattaatta	acacctttta	5640
ttgcatattt	tcgtcgttta	ttcgttgatc	gagggtgctt	ttcggctgat	gggtgcacaa	5700
attcggtaat	tgtgcatcca	tcggctgaaa	atgctccaga	atttgcgaa	gaacggtgaa	5760
aatttaagat	tttagattga	aataagccgt	tttttagaga	aaattggtcg	ttttgagaca	5820
ttaaattcaa	tttaaatccc	ctctttattt	tcagagccca	tcatcagatg	ctcaaaagcc	5880
ttccacctca	agctcagatc	tcaccgccga	gcagcttcaa	gatccaacag	ctgaagacgg	5940
caacggtgat	ggtcatgggtg	tacttgaaaa	cgtggattac	gtgaagctca	acagtcagga	6000
tagtgatgaa	cgacaacaag	agttggcgaa	tatcgcagaa	gaagcgctga	aattccagcc	6060
aaaaggatat	acacttgaga	cgacacaagt	caagacgccc	gtaccattcc	tgattcgagg	6120
acaactgaga	gaatatcaaa	tggttggatt	ggattggatg	gttacctttt	atgagaagaa	6180
tttgaatgga	attcttgccg	acgagatggg	cctgggaaag	acgattcaaa	cgatttccct	6240
gctggctcat	atggcttgta	gtgaatcgat	ttggggacca	cacttgattg	ttgtgccgac	6300
gtctgtcatt	ctgaattggg	agatggagtt	caagaaatgg	tgtccggctc	tgaagatttt	6360
gacgtatttt	ggtacggcga	aggagcgtgc	cgagaagcgg	aagggatgga	tgaagccgaa	6420
ttgtttccat	gtgtgcatca	catcatacaa	gacggttact	caagatatta	gagcttttaa	6480
gcagaggggtg	cgtagaaatt	ttgaagattt	gcggcgaaat	tggcgaaatt	gcataatttt	6540
tttaaaacca	attttaccga	taattgcgaa	atttttcaat	tttatacagt	ggtcggaaat	6600
tgctataatt	agtataattt	ttgcaaaaaat	tggtactttt	ttcgaaattt	tgaaccacca	6660
taaaacattt	ttgaacaatt	tttaagaggt	tttaataacga	aattcgttca	tttgaacaca	6720
ttttggcgat	atgaatcgcc	cgaaaatgtc	ccccaataga	cctaattttct	taacaaaaat	6780
ttaaaaaaaa	atggcccaaa	attgtctcaa	aattttcgaaa	aaaaaacggt	aattttcagct	6840
gaaatctcaa	aatttgccaa	attttccgct	tcacggagat	cagaaaaagt	tttttgcat	6900
tttttggtgt	ttatttttagc	gttatttctg	taatttagat	acatttttagc	ccaattttttg	6960
caaaaattat	actaattata	gcaatttctg	acccctgaca	aactttgaaa	ttatcggtaa	7020
acttggtata	aatggttttt	ttccaaattt	ttaaagcgat	attaaagggtg	gagtaccaca	7080
atttgaggct	ttgttttttt	ttttggaccc	aaattgggtcc	aaaactaccg	aatttctgtaa	7140
tgagacgctc	tgaaaatttc	tttctcaaaa	aaaaagttac	ggcgggttcaa	agttcgcggc	7200
aaaataaggc	ccattttcag	ctaaaaatcaa	aattttttcc	caacttctcg	gtgtctcaac	7260
gcctggaacc	taatttttat	ttattcatca	ctttttaata	aatattgtgg	tctttgattg	7320
ggctttttatt	cgttgattta	agtacattta	tggtcagtg	ggcacaaaat	gtaacttttt	7380

ttcccaaaga	ccataaatgt	actttaatca	acgaataaac	gccaatcaa	agaccacaat	7440
atztatttta	aagtaatgaa	taaataataa	ttaggttcca	gacgttgoga	caccgagaag	7500
ttggaaaatt	tttttatfff	agctgaataa	gggccttatt	gtctcaaaact	ttgaaccgcc	7560
ataactffff	tttgagaacg	tctcgttacg	aaattcggta	gttttggacc	aatttgggtc	7620
taaaaaaaca	aagtctcaaa	tttcttgfta	gagatfffft	aaaaattgat	atfffftttt	7680
tcaggcctgg	cagtacctaa	ttctcgatga	agctcaaaat	atcaaaaact	ggaagtccca	7740
acgttggcag	gctcttctga	atgtccgtgc	tcgacgtcgc	cttctcctga	ccggaactcc	7800
acttcagaac	tctctaattg	aactgtggte	gttgatgcac	tttttgatgc	caacaatatt	7860
ctcaagtcac	gatgatttca	aggattgggt	ctogaatccg	ttgacaggga	tgatgggaag	7920
aaatatggaa	ttcaatgctc	cactaatccg	acgacttcac	aaagtgtctc	gtccggttat	7980
tctgcgggcg	ctcaagaagg	aagttgagaa	gcagctgcca	gagaagactg	agcatattgt	8040
gaattgttcg	ttgtcaaagc	ggcagagata	cctgtacgat	gactttatga	gtcgtagatc	8100
aacaaaggag	aatctaaagt	ctggaaatat	gatgtcgggt	ctcaacattg	tgatgcaact	8160
ccgaaaatgt	tgtaatcacc	cgaatctctt	cgaagccggg	ccagttgttg	ctccggtcgt	8220
cgttgagaag	cttcagctcg	atgttccggc	tcgtctcttt	gaaatttcgc	agcaagatcc	8280
ctcctcctcc	tcagctagtc	aaattccgga	aattttcaat	ttatccaaaa	tcggctatca	8340
atcttccggt	cgatctgcaa	aaccactcat	cgaagagctt	gaagcaatga	gcacttatcc	8400
ggagccacga	gcaccagaag	ttggcggatt	tcggttcaat	cggacggcct	ttgttgcaaa	8460
gaatccgcat	acggaagagt	cggaggacga	aggtgttatg	agaagtcgtg	ttctggtgaa	8520
tttttaggaa	aattgagaaa	atgatctaata	tgttgaaatt	tttaaagaat	ttatgggcca	8580
caagccgatt	tgcgggaaat	tttgattttt	ggcgatttgc	cgaaaatttt	gatttttggc	8640
gatttgccag	aaattttgat	ttttggcaat	tatccgattt	gccgggaaat	ttgatttttg	8700
gcgatttgcc	agaaattttg	atfffftgga	attatccgat	ttgcccggaa	ttttgaattt	8760
tggcaatttt	ccgatttgcc	ggaaattttg	atfffftgga	atftgcccga	ttgcccggaa	8820
ttttgatttt	tggcaatttt	ccgaatttgc	ggaaattttg	atfffftggg	atftgcccga	8880
aattttgatt	tttggaattt	tgcctatttt	tcggaaattt	tgattttttg	caattttgcc	8940
atftgtcgga	aatttttgatt	tttggaattt	tgcgattttg	ccggaaattt	tgattttttg	9000
caattttccg	atftgcaaaa	aatttttgatt	ttggcgattt	tgcggaattt	ccggaaaaac	9060
atftttgtgag	ccaattttct	cgaaattttg	gcttcaatat	tttcaaatat	ttccaaattt	9120
tccactgatt	ccgaatatct	aagtaaaaaa	aaattccctg	atftttatatt	tcagcttaaa	9180
atcgctaatt	ttcgcgtcag	agacgacgtc	atgtgtcgat	ttactggatt	tttaattctt	9240
gtcggatgct	aattttccgt	tttcaacgag	tttcttccat	ttccatcggt	ttttgacgaa	9300
gttttctttg	aaaatatggt	cttaagggtca	attaaacggt	ttattatcaa	aaaaaactag	9360
caaaattggc	tttaaaaaaca	catttttcaca	gaaaactccg	acaaaaaccg	acgaaaatga	9420
aggaaacccc	ccgtttgaaa	acagaaatta	gcattctgata	aagattaaaa	tcccgtaaat	9480
cgacacatgg	cgtctggcgt	ctctggcacg	aaaagtcgog	atftttaagct	gacatacaaa	9540
aaaagaggga	tatatftttt	tacgaatttt	tcacatagat	attcgaaatc	aggggggaaa	9600
atfttgagaa	atfttgagaaa	atfttctcaga	tttccgatta	aaaatattca	atftttgttt	9660
tcttatatta	aaaaaaaatt	aactttttata	atfttttcagc	caaaaaccaat	taatggaaca	9720
gctcaaccac	ttcaaaatgg	aaattcaata	ccacaaaatg	ctccaaatcg	tccacaaact	9780
tcatgcattc	gttcaaaaac	cgtcgtaaat	acagttccac	tgaccatctc	caccgatcga	9840
agtggttttc	atftttaatat	ggccaatggt	ggaagaggtg	ttgttcgttt	ggatgattca	9900
gcacgtatga	gcccaccgct	caaacgtcag	aagctcaccg	gaactgcaac	gaattggagt	9960
gattatgttc	cgcgacacgt	tgttgaaaag	atggaagaat	cgagaaaaaa	ccagctggaa	10020
attgttcgaa	ggcgatttga	gatgattcgt	gctccgatta	ttccactgga	aatggttgog	10080
ctggttcgag	aggaaattat	tgcagaattt	ccagttttgg	ctgtggaaga	ggacgaggtt	10140
gtgcaggaga	ggcttttgga	gtattgagag	ttgttggtgc	aaaggtagaa	ttttgaaaat	10200
tattactttg	ctttttttta	aaccaaaatt	ggcccaaaac	taccgaattt	cgtaattgaga	10260
catttctgaaa	gcttctcaaa	aaaaaagtft	tggccgctca	aagttcggga	aaataaggcc	10320
catttttcagc	tgaaatcaaa	atftttttcca	acttctcggt	gtcgcaacgt	ctggaactaa	10380
aatttttgaa	aacgagaaat	tttccatttt	ttgcaagctg	aaaaatcaaa	gttttttttt	10440
cctcaaaaatt	ggacaaacaa	aaaaatfttt	ttttgaaaat	tgatcgaaaa	aattcaaaat	10500
ttctataatt	tttcgatttt	ttaaataaaa	ctttcatcat	ttttcttcca	aatttagttt	10560
tctcgatttt	aactfttttt	aaaaaaaatt	tttttaatac	gaaaaaaatt	caatttttag	10620
tctaattctt	tttttagacc	aaattgggtc	aaaactaccg	aatttcgtaa	tgagacgttc	10680
tgaacatttc	tcaaaaaaaa	gttatgacgg	ttcaaagttc	ggcaaaataa	ggcccaattt	10740
catataaaat	caaattfttt	ttctaacttc	tcggtgtcac	aacgtctgga	acttaatttt	10800
tatttaatta	ttacttttca	ataaatattg	tggtctttta	ttaggcggtt	atftgttgat	10860

ttaagtacat	ttatggtcaa	gtggggccca	aataaaaagtt	acattttgtg	cccacatgac	10920
cataaatgta	cttaaatcaa	cgaataaacg	cctaatacaa	ggccacaata	tttattaaaa	10980
agtgttgaat	aaataaaaaat	taggttccag	acatttgtgac	accgagaagt	taaaaaaaat	11040
tttgatttta	gctgaaaatg	ggccttattt	tgtctgaactt	taaaccgcta	taactttttt	11100
tttgagaatt	ttcagaacgt	ctcattacga	aatttcggtag	ttttggacca	atttgggtct	11160
aaaaagaat	tagagctaaa	attgaatttt	cttcgtatta	aaaatttttt	ttttgaaaaa	11220
agtaaaaatc	gagaaaacta	aatttggag	aaaaatgatg	aaaattttat	ttaaaaaatc	11280
gaaaaattat	agaaattttg	atcgattttt	tcatcaatt	ttcaataaaa	aattttttgt	11340
ttgtccaatt	ttgaggaaaa	aaaaaacttt	gatttttcag	cttacaataa	atggaaagtt	11400
tctcgttttc	caattttttg	atgtggattt	ttatgagaaa	aaatatataa	tgtcacaaaa	11460
aatagattat	tatctaaaaa	tcgaaaaaat	taaattttcc	agttttcagg	aaaaaaatcg	11520
ttaagaaatt	gtttttccat	taaaggtgga	gtaccgaatt	ttgagacgct	gcttttttag	11580
acccaaaatg	gtccaaaact	accgaatttc	gtaatgatac	gctctgaaaa	attttcaaaa	11640
aaaaagttgt	gaccgctcaa	agttttggaa	aaatggcata	tttttagcta	aaatctcaa	11700
ttttggcaac	ttatcgggtg	cgcagcgggt	ggaacttaat	ttttatttaa	ttgtcattca	11760
ttaatgcag	ttttggcatt	tcattatgtg	ttatttcgtt	gattgagatg	ctttttgtgc	11820
ctgcatcgac	caaaaaacca	tctcaatcaa	cgaataaaca	cataataaaa	tgccaaaaa	11880
tgcattaaag	gatgataatc	aaataaaaaat	taagtttcaa	ccgctgcgac	accgctaagt	11940
tgccaaaatt	tgagatttta	gctaaaaatg	gtccattttt	ctaaaaactt	gagcgggtcac	12000
aacttttttt	ttgagaaatt	ttcagagcgt	ctcattacga	aaattggtag	gttcggacca	12060
atttgggtct	aaaaaagcag	cgtctcaaaa	ttcgggtact	cacctttaaa	gttttcaatt	12120
taaagtataa	attatccaat	caaaaaattga	cgaaaaaatt	ttttaaaaat	ttttcttcc	12180
gaaaaaaaaa	ttaattttta	tttttgttag	attcggaaatg	tacgtcgaac	cagtgtctgac	12240
cgatgcttgg	cagtgtcgtc	catcatcgtc	tggtcttcca	tcatatattc	gcaacaattt	12300
atcaaatatc	gatctgaatt	ctcgttctct	tctcctcaac	acctccacta	atttcgatac	12360
ccgaatgtcg	atctcacgtg	ctcttcaatt	cccagaactc	cgtctgatcg	agtacgattg	12420
tggaagctt	cagacgttgg	ctgttctgct	tgcgtcagttg	tacctgtaca	agcacagatg	12480
tctgatcttc	acgcaaattg	caaagatgct	cgacgttctg	cagaccttcc	tttctcatca	12540
cggttatcag	tatttccgcc	tcgacggtac	cactggtgtc	gaacaaagac	aggcgatgat	12600
ggagcgggtc	aacgcggatc	ccaagggtgt	ttgcttcatt	ctgtcgacga	gatccgggtg	12660
tgttgagtc	aatctaaccg	gtgctgacac	tgtgatcttc	tacgattcgg	attggaatcc	12720
gacgatggat	gctcaggctc	aggatagatg	tcatcgtatc	ggacagacga	ggaatgtctc	12780
gatttatcga	ttgattttccg	agcgaacaat	tgaggagaat	attctgagaa	aggcaacaca	12840
gaagcggcga	cttgagagat	tggcaattga	cgaggctggc	ttcacaccgc	agttcttcaa	12900
acaatctgac	agtattcggg	atctttttga	tggagagaat	gtggaagtga	ctgctgtggc	12960
agatgttgcg	acgacgatga	gcgagaaaga	aattggaggtt	gcgatggcaa	agtgtgaaga	13020
tgaagctgat	gtgaatgcgg	cgaagattgc	ggtggccgag	gcgaacgttg	ataatgcgga	13080
gtttgatgag	aaatcattgc	cgccgatgag	caatttgcaa	ggagatgagg	aggctgatga	13140
gaagtatatg	gagttgatac	aacaggtaaa	attcggcgga	aatcggaat	tttcccattt	13200
agaatatcaa	attttgcccg	attgtgtcgt	tttttgattt	ttcgatttat	tcgatttggt	13260
tttgagggaa	aatcggaana	atgttcagaa	aattaaccat	aacatgtgat	ctttttaaaa	13320
tcttagcgca	aatgtcttct	aaaaaataaa	gaatgaccaa	aaatttttaag	ctaattttttg	13380
aaaaaccaaa	gaaaaaattt	agatttttctg	atgttttccg	agacaaaaag	acaaaaacgg	13440
aaattgtcga	aaatgaatga	aatttttaatt	ttttcagcaa	aaaaaaaata	gtacttaatt	13500
ttaaaaaatg	tgatcatttc	ggtaggaaaa	tctggaaaaa	tcgattttca	aacaaaaaaa	13560
aaccgagcct	ctacaatctt	tttttttccc	gaaatctcca	gaacttctca	caataacaac	13620
tatataaatt	tcaaaatttc	agctcaaacc	aatcgaacga	tatgccatta	actttcttga	13680
gacacagtac	aagccagaat	ttgaggaaga	atgcaaagag	gcagaggtat	attattccat	13740
tcatctgact	tttttttttt	tttttttaatt	ttaaatttca	ccaaattaat	tacaggctct	13800
tatcgaccaa	aaacgcgaag	aatgggacaa	aatctcaac	gataccgccg	tcattgacct	13860
cgacgattcg	gatagtctgc	tgctcaacga	tccttcgact	tctgccgatt	tttatcagag	13920
ctcaagtctt	ttagacgagg	tagcgatcg	tcgtcgtcgc	agcagcagcc	ttctccaaaa	13980
agccgctcaa	aaaccggcaa	aaaagcctca	aaacttccaa	attcgtgtct	gctccccgtc	14040
taagcgtaaa	tctcaggctc	cttccttcga	tccatatgtt	tcgtacgcac	cgcacgcgtc	14100
cgcttctccc	ccggattccc	cgcgtaaagag	aagatcacgt	ggtgcgcgta	gtttaggtag	14160
tggtgggtgg	ggtgggtggg	gtagtagatc	tggtggaaga	cctgcccgcc	gatcagtgaa	14220
gaaagaagaa	tcagatgatg	atgatgagga	ttattgccaa	gaagaggaag	tgaagcgaaa	14280
tccggcagaa	aaggtcccgc	cgaaaagaaa	acgagttgtg	tttgtggaac	ctccagaggt	14340

gaagccgccc	gagccgaaaa	aacgagttgt	tgttcctgct	ccatcatcat	catcatcagc	14400
tctaactact	cttccacaac	aaggaccgct	gatttcggtg	ccaaaagctg	tgccagttgt	14460
acctcgcccc	caacaacaag	caccaccaca	gctcatcaaa	aagcaccagc	agactctgat	14520
gcctgtgaag	gtgctcaaga	ttagtgggtg	tggtgggtgg	actccaggac	catccagttg	14580
atcgccaggt	ccatcaatcc	tccgaagaac	cgttgttcca	ggcataggcg	ctgggtgggt	14640
tggacgccta	ccgcttgtca	gaatgcctgt	tcgccctcca	tttcctggct	cgcaagctcc	14700
tgctccaccg	ctgagaagtg	gtgttgctcc	aacagctcct	gcagcagctc	cagccagttt	14760
cgctgttccg	tcgtcgagag	ttcgagttat	cacgacgaga	actccggtcg	ccaccaccat	14820
ggtgcaacaa	caacaaagcc	cgagcccgtt	gatgtttcca	gtccgggttg	tgcaaaaggcc	14880
cgggccatct	ggaccaccac	cacctggacc	tccagatcgc	ccaggatttg	gaatctatga	14940
gaagccgaga	ttctcacttg	gatcacgaag	aagccgtgga	gattcggggc	cggaaagatcc	15000
ggcgccacca	cagccaccac	caccaccac	ttctaggcca	ccgccacaag	cctaggcgct	15060
aggattttcc	tttttttttt	gttgattttt	gctctttttt	tgctctctca	tgattttata	15120
atctcatttt	gctttaatat	ttccattttt	ttggatgtgt	ggaatttttt	tttttgaaaa	15180
tcgggaaaaa	acgaaaaatt	tgaacttttt	ggtgattttc	agagaaaaat	ccgtttttta	15240
atgaaaaaat	cggataaatt	cagatttttc	gaaaaaaaaa	accgagaaaa	tttcaaattt	15300
tcagtttttt	ttttcaaaaa	atcgaaaaaa	aaagtaaatt	ttcagaatta	tcagccaagt	15360
ttttgcgatt	ttttgaaaaa	tttcaatttt	tggaattttt	tggaaaaaaa	tcaattttta	15420
attcgaaaaa	ttggaaaaat	taagattttt	cgaaaaaaaa	aacgaagaaa	gtttcaaatt	15480
tttagctttt	ttcaaaaaat	cgaaaaatcg	aattttttta	atttttcgaa	taaaaaaaat	15540
cgaagaaatt	ccaaaacttt	gcgttttttc	ttgaaattat	ctgaaaaccg	gaattttttt	15600
tcaaaattcg	ccattttttg	cgaatttttg	taatcttttt	ccgagaaaaa	tcgatttttt	15660
aaatcttaat	aattcagatt	tttcgatttt	cttttggtcc	aaaaagtcaa	aaaccgaaca	15720
attattttat	tcaaaaactc	taaaaatttt	caattttttg	gaaatttttc	ggtataaaaa	15780
aaacccattt	ttaaatcaaa	aaatcgaaaa	tttttgtgat	ttttcgattt	ttttcactcc	15840
aaaaaaaatt	cacacagcaa	aaaataaaat	ccgcgcattt	ttgagcgcac	ctttcaatgt	15900
tttaattctt	atcacgacgt	caaaattcgg	ttatttttca	cacacacaca	ttttctccc	15960
gagcggttct	tttttctatg	agttctccca	tgttttgttt	ttatatttga	gacatttttt	16020
tttgttgata	agtttcaact	tcttcttctt	cttctgacta	taaacgtttt	tctccatgtt	16080
ttttgcctgt	tttctgccga	ttttttgaca	cccaaaattt	tttttcattt	tcgctcgaaa	16140
atgcacgtcg	ttggctctag	ctttggcaag	tttttaacac	tgattttctg	gttttttttt	16200
ttttttgcag	aattttttcag	agataggggg	ctcattccag	cagggtttcc	cactatattt	16260
cgcatttttt	ccaaaaattt	ttgtattttc	aaaaatttcc	aaaaagaaa	gggttttctt	16320
taccaaaatt	ttctcgccac	ttttggctta	attttggtt	tagagattcg	atcgaaaaaa	16380
ttgcgaaagt	ggcgagaaat	ctcactggtt	tgatgtttga	ccccctacta	tagaaaattt	16440
gaaaaaaaaa	aaaaaaaaaa	aaaactagac	gaaatttggt	gaaatcttgc	tggaatttga	16500
cgagtcgatg	gtggattttt	cttgaaacga	atgaaacggt	gatttttggt	cggagaaata	16560
tggcgaaaaa	tggtgagaaa	tgacgaggag	gaggaagaag	ctgaaaatct	ggaggaacaa	16620
aaattgtgtg	gaagtctcgg	gaagaaatta	gaattgaaat	tttaaagtgt	tctgagaatt	16680
ttttgtgtga	aattttttta	aatctgtaga	tcaaatatca	aaaaaaaaaa	tcagaactat	16740
tacgtgttta	tccacaaaga	tgagaaaaat	cgccatatct	ggcgcgcaaa	tgaacccgcg	16800
ggaagagaca	aaactactgt	agtttttaac	caatttggtg	agatttacga	gctattgcgt	16860
catcgaattg	aattttaattt	tcaggcggtt	cacacgtttt	tatattgaaa	tttatctatt	16920
tattgaatca	atcttaaaa	aaaacacaaa	aaattttttt	taaaaattgc	ggctcaaaat	16980
taaattcaat	tcgatgacgc	aatagctcgt	aaatctacac	aaattggtta	aaaactacag	17040
tagttttgtc	tcttcccgcg	ggttcatttg	cgcgccagat	atggtgattt	ttctcatctc	17100
tggtataaca	cgtaataaca	tttctcggca	caataaaatt	ttgctgaaac	aagtgcgcgc	17160
ctttgaagag	tactgcaatt	tcaaacacgg	ttttttggtt	ggaaagcaca	gtactttttc	17220
aaaggtgcac	accttctcga	atttctcttc	gtgtcgagac	caagaatgcc	atttttcgat	17280
ttttaaaaaa	tcaaaaaaaa	aattaccttt	ttaaaggtgg	agtaccgaaa	tttgagactt	17340
tgtttttttc	ggcccaaaat	ggtccaaaac	taccgaattt	cgtaatgaga	cgttctgaaa	17400
atttctcaaa	aaacaacgtt	atggcggttt	aaagtccagc	aaaataaggc	ccattttcag	17460
ctaaaatcaa	aatttttttc	cagcttctcg	gtgtcacaa	gcctggaacc	taatttttat	17520
ttattcatca	ctttttgata	aatattgtgg	tcttttatta	ggcgtttatt	ttattgattt	17580
aagcttattt	atggtctttg	tggcgttaca	ttttgtacct	taaaaacat	aatgtacttt	17640
aaatcaacga	ataaacgcct	aatcaaaagg	tacaatatat	agtagaaagt	gataaaaaat	17700
taaaaattga	gttccagacg	ttgcgacacc	gagaagtgtg	cgaaaacttt	gatttttagct	17760
aaaaataaag	catttttcca	aaactttgag	cggtcataac	ttttttttga	gaaagaaatt	17820

ttcagaatgt	ctcattacga	aattcggtag	ctttggggcca	ttttggggccg	aaaaagcaaa	17880
gtctcaaat	tcagcactcc	aacttttagcc	tttaccttgg	tgaaattttt	taatctgtag	17940
tatactttat	ttttggccga	ctttttgaac	acaaattcgg	tgtagttta	aaaaacaat	18000
caaaactaac	atattatcca	gacgcgaaat	ttttgtcgg	tttcttcgcg	ccaaaaagta	18060
cggtaacagg	tttcggcacg	atacattttt	gttaaaagg	gctgctcctt	tgaagagtgt	18120
ctaataat	tcaactttcg	tttctgttgg	aattttcttc	aatttttcat	agatgttttc	18180
gatgaaacaa	aaaattaaca	caaaatcgtc	gtgtcgagac	cggaaaaaat	tttgcgtctg	18240
tgcaacaaac	ccggaaaatt	aaagtagcat	attgatccaa	attgccgatt	tgccggaaat	18300
tttgattttc	ggcaatatac	cgatttgccg	gaacatttga	ttttctggaa	tataccgatt	18360
tgccggaatt	tttgggtttt	ggaaatttgc	cggaaattta	gaattccggc	aatatgccga	18420
tttgccggaa	attttgattt	tcggcaatat	gccgatttgc	cggaaatttt	gattttcggc	18480
aatataccga	tttgccggaa	catttgattt	ccggcaatat	gccgatttgc	cggaaatttt	18540
gattttccggc	aatatgccga	tttgccggaa	attttgattt	tcggcaatat	accgatttgc	18600
cggaaacattt	ggttttccggc	aatatgccga	tttgccggaa	tttttggttt	tcggaaattt	18660
gccggaaatt	tagaattccg	gcaatttgcc	gatttgccgg	aaattttgat	ttccggcaat	18720
atgccgattt	gccggaaaatt	ttggttttcg	gaaatttgcc	ggaaatttag	aattccggca	18780
atatgccgat	ttgccggaaa	ttttgatttc	cggcaatatg	ccgatttgct	agaagaaatc	18840
gtttgtcacc	cacacgtgta	ttgatttgat	ttttctagat	aaaattctac	gacgagctgg	18900
acgatcat	gccaatctgg	cttcaccat	caccaccaga	ttcggatgcg	gatttcgact	18960
tgagaatgga	agatgattgt	ctcgatctga	tgtatgaaat	tgaacaaatg	aacgagctc	19020
gcctaccaca	agtttgtcat	gaaatgagac	gtccgttggc	tgaaaaacag	cagaaacaga	19080
acacgttgaa	tgcgtttaat	ggtaatat	tcaaaaaaaa	atttttttga	aaaaattcaa	19140
ttaaattcga	ttttgagcaa	tttttatcgt	gaagattgca	taattttgag	attttgcgcc	19200
aagatttttg	ttaaattgaa	aaaaagagat	gtgcgccttt	atggagtact	gtagttttga	19260
aaattgaaat	tacagtactc	tgtttaaagg	cgcacacatg	tattacgtag	cgaagaaaa	19320
agtacagtaa	ttagttaa	aagactactg	tagcgttgt	gtcgatttac	gggctctgaa	19380
ttttatatga	atttttgaaa	actagaaaca	tctcaaatg	cataaaatta	ccatttgaa	19440
ctcccgccaa	gtgattttgt	tcgacggggc	gcgcttgac	gttttctatt	ttaattta	19500
tcaatttttt	ttgcttaatt	ctcaccgat	tttcatgttt	tcagtttgat	tttgatgaa	19560
atttgaggag	aatatcaaca	taaagtcttt	tcaatcgaaa	atgtgcattt	atattgacat	19620
tttctccgaa	tttccatcaa	aattaaactg	aaaacacgaa	aaatcggatg	gaattaagcg	19680
aaaaaattga	gttaaatgaa	aatagaaaac	gtgcaagcgc	gctccatcga	acaaaatcaa	19740
ttggcgggag	gttcaaatgg	gaattgtatg	caattttcaa	aaggctcgat	aaaattttga	19800
agaaagcaaa	ttaaatttaa	aaaatcgagc	tcgtaaatcg	acacaggcgc	taattttcaa	19860
aaaaataaaa	tgacacccaa	aaaatcataa	gaaaatcata	aataaatatt	acgggaacac	19920
aaaactcaga	gaaccctgat	tgcaacaacat	atttgacgcg	caaaatatga	aatatctcgt	19980
agcgaagaa	aaactaccgt	aatttaaaaa	catttaaatg	actactgtag	cgcttgtgtc	20040
gatttacgag	atctcgattt	tctaaataaa	ttttttaaaa	aatgatgtca	gcgatattcc	20100
atttgacttt	gtttcttcgt	attattttct	catttttgct	tgattttatt	taattttata	20160
attttattta	aaatcaagca	aaaacgagaa	aataatacga	agaaacggag	ttaaatggaa	20220
tatcgctgac	ataattttaa	aaaaaaattt	aattagaaaa	tcgagatccc	gtaaatcgac	20280
acaagtagtc	atagtagagt	agtcatttaa	ctaattactg	tacttttctt	ttcgctgcga	20340
gatatttcat	attttttatt	atatttttat	ttattttcat	atttttatat	atatatatat	20400
atatatat	cttggcggtt	taatgcagtt	tctctcaatt	aattccagac	attctatcgg	20460
caaaagaaaa	ggaatcgggtg	tacgatgcgg	tcaacaagtg	ccttcaaatg	ccacaatccg	20520
aagcgatcac	agcagaatct	gcagcgtctc	cagcatacac	ggaacactca	tcattctcga	20580
tgatgatac	aagccaggat	gcgaagattg	agccaagttt	gactgaaaat	caacaaccca	20640
ccaccaccgc	cactactact	actacagtac	ccaacaacag	acaacaacag	cagcagcaaa	20700
aatcgctcgaa	aaagaagaga	aatgataatc	gaacggtacg	gaggttacta	gcgaacaatt	20760
tcaagaaatt	ttgaatttgt	gaaaattcaa	ttccggcaat	tttctgattt	gccggaactt	20820
ttaattttcg	ccgaattgtc	aatttgccgg	aaattttgat	ttccgccgaa	ttgtcgattt	20880
gccggaactt	ttcattttcg	gcaaattttc	gatttgccgg	aactttta	ttttgacaaa	20940
ttgtcgatgt	gccggaaatt	ttgattttcg	acaatttgct	gatttgccgg	aaatttcaat	21000
ccaacaatt	ttccgatttg	ccggaaattt	caatcccaac	aattttccga	tttgccggaa	21060
atttcaatcc	caacaatttt	ccgatttgcc	ggaaatttca	atcccaacaa	ttttccgatt	21120
tgccggaaat	ttcaatccca	gcaattttcc	gatttgccgg	aaatttcaat	tcgggcaatt	21180
tttcgatttg	ccggaaactt	tcattttcgg	caaagtgtcg	atttgccgga	acttttcatt	21240
ttccgccgaat	tgctgatttg	cccgaacttt	taatttttga	caaattgtcg	ttttgctgga	21300

aatttttgatt	ttcgacaatt	tgccaatttg	cgggaacttt	taatttttga	caaattgtcg	21360
atttgccgga	aatttttgatt	ttcgacaatt	tgccaatttg	cgggaacttt	tcattttttgc	21420
caaattgtcg	atttgccgga	aatttttaatt	cgggaacttt	tgcgattttgc	cggaaaatttc	21480
aattccggga	atttaaaaaac	actaaaaaac	aaaaattttc	ggttttcccg	tttttcgatg	21540
tttcagcttt	tctcaaaaaa	ttgcgattcc	cggaaaaatc	gaaacaattt	tcgggggttaa	21600
aaccgggaaa	ttcctaaatt	cctattttaa	agaattgaâa	aaaaactctc	aaaattccag	21660
gctcaaaatc	gaacagctga	aaatgggtgtg	aaacgagcga	caactccacc	accatcatgg	21720
cgtgaagagc	cagattatga	tggagccgaa	tggaaatag	ttgaagatta	tgcactactt	21780
caagcagttc	aagtcgaatt	tgcaaagtct	catttagtcg	aaaaatcggc	gaatgagggg	21840
atgggtgtga	actgggaatt	cgtgtcgaat	gccgttaata	agcagacaag	atttttccgc	21900
tcggcccgtc	aatgctcaat	tcgatataca	atgtttgttc	ggccaaaaga	gctcggacag	21960
ttgggtggctt	ctgatccgat	ttccaagaaa	acgatgaaag	tcgacctatc	gcatactgaa	22020
ttatctcatt	tgagaaaagg	acgaatgact	acggagagcc	aatatgctca	tgattatgga	22080
atattgactg	ataagaaaca	tgtgaataga	tttaaaagtg	ttcgagtggc	ggcaacacgg	22140
agacctgttc	agttttggag	aggccctaaa	ggtagaggag	gatggcttca	taatagtac	22200
tgcaactttt	ttctcacgag	ggacgagaaa	aagtgggttc	tagggcatgg	ccgaggtgcc	22260
gacaagtttc	agcggccatt	tatcttgctt	tgttttccgc	ccgttttctt	tcgtttttca	22320
ccgatttttt	tcgttttttc	ttataaaaa	tgataaataa	atattttttg	cagatgctaa	22380
aaaaatttcc	aagtaaaaaa	atcatgtatt	cagtgggcat	gcagcgggtg	aagtgggcat	22440
tgtaatatga	tggattacgg	gtatacaaaa	cctaaacttt	ttctgaaaca	tgatacatgt	22500
gctgcttaaa	tgctgagact	acctgatttt	cataacgaga	ccgctgaaaa	agttttgagg	22560
tttccaaaat	tcaacttttt	tggtgaaaaa	gtcgagattt	tcgcacaaaa	agttgaattt	22620
tgaaaacctc	aaaacttttt	cagcgggtctc	gttatgaaaa	tcaggtaatt	tcagcatcta	22680
agcatcata	gtatcatgtt	tcagaaaagt	ttaggttttg	tattcccgtg	atccatctat	22740
ttacattgac	cacttttcacc	gctgcttgcc	cactgaatac	ataatttttt	cacttggaâa	22800
ttgttttagc	atctggaaaa	agtattttatt	tatcagtttt	aataagaaaa	aacgggaaaa	22860
agctgtgaaa	aacaaaagaa	aacaggcgga	aaacaaagca	agataaatgg	ccgctgaaac	22920
ttgtcggccc	ctcggccatg	gcctagaaac	cacttttctt	cgtccctcgt	gaggaaaaag	22980
ttgcagtgat	agtctaaaat	tcggaggaat	tttttaaaat	tggaââââat	tgtaaaattt	23040
ttttttctcg	gaaattggaa	aatcacaaat	tttcgatttt	tgtttggtta	ââââââââag	23100
aaaattggca	taataaaaca	tttctttttt	ttttgaaâat	tgggaaacttc	ttaatatcag	23160
attttttaag	taagattttt	ttgattttcc	ggaaattcgg	aaaacctgaa	aatttttcaac	23220
atttcaaaat	ââââââââââ	gttttttttt	tctgaaââââ	tccaacaaaa	aaaggtcaaa	23280
tcgtcagaat	tattgttgga	agtggcgggt	tttcacgatt	agagttcagt	attttttctt	23340
ctgaatttca	aatttgaaaa	ââââââââââ	aaactgtaga	ââââââââââ	ââââââââââ	23400
ââââââââââ	ttaaaggtaa	agggaâââââ	gaccgtaatg	accgaatata	actgttgaaa	23460
atatcaacaa	ââââââââââ	gaattttttg	tgactttttc	aattttttcaa	gaataââââââ	23520
aacgaccgaa	tââââââââââ	gaattcccgc	gcaaââââââ	gactgggttct	ggccaaattta	23580
cagtcttttt	ataâââââââ	ââââââââââ	ââââââââââ	attttagccag	ââââââââââ	23640
ââââââââââ	tgacgtcact	catttgccgcg	cggâââââââ	atttââââââ	gccgttttctt	23700
tgattttttga	ââââââââââ	ââââââââââ	ââââââââââ	aattttttttg	aatttttttac	23760
agtttttttat	tcgggtcatta	tgggggttatt	caagtagtgt	cggâââââââ	ââââââââââ	23820
ââââââââââ	tcacaactct	gtattcaagt	ataâââââââ	atgtatttaa	atacattttg	23880
ctacattact	tgaataââââ	cattaggggt	tattttcttt	agagcaââââ	ââââââââââ	23940
tggtctctact	ccacctttta	atgâââââââ	cgacaatttg	tgatttttgca	atttccagaa	24000
ââââââââââ	ââââââââââ	tttggaââââ	accâââââââ	gccatttgaa	ââââââââââ	24060
ttccââââââ	aattattttg	cagctctaga	atctcgââââ	ctgcaatctc	tââââââââââ	24120
aatgccacca	cgacacgagt	cgagactcgc	cgaattcgac	gtââââââââ	atattcgcct	24180
ggacgccgag	gacattgtca	caatgtccga	cgagtcgatt	gtcgcctatg	aagcagagcaa	24240
gaagaagcta	ctggccagtc	gtcaââââââ	accctcacca	cgtcaagatg	tccgattcca	24300
tacgttggtt	cttcggccgt	ataccgtacc	tgtgacaact	gagtactcgg	ctgcaccttc	24360
tcgtcgtgaa	atgcgcacgc	ctgttccacc	gcttcagcct	tcggctttat	tcacgatttc	24420
ctcagttgct	gctgctgcca	cgtctggggc	actaccatca	attcagcatt	tcagtcgtc	24480
gtctacgggc	ttgggatctc	agcaââââââ	gcaâââââââ	cataattctg	agcaââââââ	24540
taatgtgcaa	aatatgcac	ââââââââââ	taattcaagt	caââââââââ	caatacctat	24600
ccgacaaatc	ggagcagcat	catcacacca	acatgatcaa	ggatctcagg	ggcctggggg	24660
ââââââââââ	gcctatcacc	tggtgcaaca	gggatcacag	caacagcagc	agcagcagca	24720
gcaggcagacg	ttacagcgaa	gaaatgcggc	ggcggcggga	gggtcgaatg	tcagttttat	24780

tcagcagcag	cagcagcagc	agcaatcggg	taaaaattgt	atggatttat	aggaaattat	24840
atgaatttgc	gcggggatag	ccccggcgaa	aaacgggaaa	aagcgacaat	ttaaaaaaaa	24900
atcgtgtgaa	aatctcaatt	ttttacaatt	ttgaaagtaa	ttttttattg	aaaaaagtgg	24960
aatttaggca	ttcatccaga	gcagggtctg	gaccaaiaaaa	aatttttggg	ccaaaaacca	25020
aaaaacaaaa	aattgaaatt	tccgaaaaat	caacttaagc	atcaaaaaatt	ttttgttttt	25080
ttttttgttt	tttggttttt	tttgggtatt	tgacgaaaaa	acgatttttt	ggttttttgg	25140
tttttcgaga	ccaaaaaaac	caaaaaaatcc	aaaaaaatgt	ttgccgtgtc	tagtctcgac	25200
ctagacacgg	caaacatttt	ttttttttgg	attttttggt	ttttttggtc	ccgaaaaacc	25260
aaaaaaacca	aaaaatcgat	ttttcgtcaa	aataccaaaa	aaaaacaaag	aattcccagc	25320
ccctttcgcc	aaaattgccg	gatattttca	aacctcaaaa	aaaattttata	aaggtggact	25380
acatcctgtg	gggaaattgc	tttaaaacat	gcctatgggc	tcacaatgac	cgaatatcat	25440
gattaaaaaa	ttcaacaaaa	aaattactag	atttttatgt	attttttgaa	aattaaaaaa	25500
atctcagttt	tcaacctaat	tcctatttga	atttcgcgca	atttgatttg	ttcgatggag	25560
cgcgcttgct	ttattttttt	ttattcattg	attttatttt	tattagcatt	atttcactga	25620
ttttcttcat	tttttgtgtg	tttttgggtg	gaattgaaat	gaaaaaaaac	aagataaatg	25680
cagaaagtct	gttaaaaagg	cattgaaaaa	gcttaaaacg	gcaacaagct	tgaattttgt	25740
atatttttaca	cagttttacg	catttttcaat	gacttttttaa	caaactttcc	gcatttatct	25800
tgtttttttc	agttcaattt	ccattaaaaa	acacacaaaa	aaaaatgaag	aaaatcagtg	25860
aaataagggt	aataaataaa	ataaatgaat	aaaaatgatg	caagcgcgct	ccaacgaacg	25920
aattcaattg	gcggaaattc	aaatatggaa	ttaggtgaaa	actgagattt	ttttttcaat	25980
tttcaaaaaa	tcatataaaa	tctagaacca	ttttttgaat	tttttaataca	tgatattcgg	26040
tcattgtgac	cccataggcg	tgtttttaaag	caattttcccc	acagggtgta	gtccaccttt	26100
gacgaggttt	gaaaatgtcc	ggcaattttg	ccgaaattgc	cggaaacttg	agattttttca	26160
gtgaaaaatt	ccaaattttca	tgtggaaaac	tgtttttttg	ttttttggaa	aatgcaacaa	26220
aaaaaaactat	ttggcgcgaa	aacgcggata	gtttttgcaa	ttttcaagga	ttttccgcta	26280
tttttaattgt	ttttatgccc	aatttttact	taaaaaatca	taattattcg	gaaaattgct	26340
gaaggagctt	tccaattgtc	tgtggagcgc	gtttgactaa	tcagataata	ttccaggcgg	26400
tcaaggacaa	agcttcgttg	tcatgggtc	gcagagctca	tcaaatgatg	gacaagggtg	26460
agcatcgacc	gtcggaggag	gaggaggagg	atcacaacag	cctcaccagc	agcagcagca	26520
gcagccacaa	caaagaatac	agtacattcc	acaagttacc	ggtagcggaa	ataacggtgg	26580
aggtggtgga	agaggaggct	acggtagtac	actggtcatg	ccaagaggag	gacgtgttgt	26640
cagggtttggt	agaaatacaa	aatcgcgaaa	aaacggcatt	tccggcttcc	cgaccaatca	26700
gcgatttgct	ccgcccactt	tccgaccaat	ccgctgaccg	aggcatttga	ttggttttga	26760
attgggcgga	gcagcgaatt	gctgatgcga	aatacgggaa	gttctcattt	tgatggaaat	26820
tctgcaaaat	tctttaaaaa	aaacaaaaat	ttctcaaatt	cggaaaaaat	cacaaaggaa	26880
atcgaagaaa	atcgcgattt	ttgattcccc	gaccaatcag	cgatttgctc	cgcgccactt	26940
tgaaccaatc	agcgttcgag	gcatttgatt	ggttcaaaaac	tgggcggagc	agcgagttgc	27000
tgattggatt	tttcagtttt	ttaaattttta	aagctttttt	taacggaaaa	attcgagaaa	27060
accatagatt	ttgatgagaa	atgatgaaaa	ttttcatgaa	aaaatggaaa	aatgattgga	27120
aattaatcaa	aaaatcttga	aaaaaaaattt	tttttcagag	aaaatgcttc	attttttggct	27180
ctgaaacgcc	tcttttttat	ttgtgcctcc	ccgaccaatc	agcaatttgc	tccgcccact	27240
tttgaaccaa	tcagcgaccg	agcgatccga	ttggttttga	attgggcgga	gctaaaaatga	27300
ttttaaaaaa	attcccgat	tgtttaattct	agaaatttag	aaaaaagaaa	tatagaaaaa	27360
aaatagaaaa	aaattaaaaa	aaaaaaaaca	aaaaatcgga	aaacgtcgga	aaatattacg	27420
aaaaaaattt	ttttaattga	ttttttttcg	aaaaaaacta	aaatttttaac	caaaaattca	27480
aagaaaaaat	ttgtttttga	tttttttttc	gaaaaaaaaa	aaaattttta	ccaaaaattc	27540
aaaaaaaaaa	tgtttttctt	gatttttttc	caaaaaaaact	aaaattttga	ccaaaaattc	27600
agcaaaaaaa	aaattttttta	attgattttt	ttttcgaaaa	aaaataaaat	tttaaccaa	27660
aattcaaaaa	aaaaattttt	tattgacttt	tttcgaaaaa	aactcaaatt	tttaaccaa	27720
attcaaaaaa	aaaaattttt	tttttgattt	tttcgaaaaa	aaactaaaat	tttaaccaa	27780
aattcaaaaa	aaaaatgttt	ttcttgattt	ttttccaaaa	aaactaaaat	tttgaccaaa	27840
aattcagcaa	aaaaaaaatt	tttttaattga	tttttttttc	gaaaaaaaat	aaaattttta	27900
ccaaaaattc	aaaaaaaata	tttttttattg	actttttttc	aaaaaaaact	aaattttta	27960
caaaaattca	aaaaaaaata	tttttttttt	gatttttttc	gaaaaaaaact	aaaattttta	28020
ccaaaatttc	aaaaaaaata	ttttttattga	ttttttttcc	aaaaaaactaa	aaatttgacc	28080
aaaaattcag	caaaaaaaa	attttttta	tgattttttt	tcgaaaaaaa	ctaaaaattt	28140
gaccaaaaat	tcaacaaaaa	aaaaattttt	tattgatttt	tttcgaaaaa	aactaaaatt	28200
ttgaccaaaa	attcaacaaa	aaaaattttt	tccagccagc	gggaactcta	ccaggcgggtg	28260

gacgtctgta	tgtcgatcat	aaccgtcatc	catatccaat	gtcgtcaaat	gttgtgccag	28320
tacgtgttct	accagccacg	caacaaggac	aacaacgaat	gatgacagga	caacgtcgtc	28380
cggctccagc	gcccgggtact	gtcgccgcaa	tggtgttgcc	gaatcgagga	gctgggtgaa	28440
ttccgcaaat	gcgcagtttg	cagtgaagttt	tgacgcgaaa	ttggacgatt	ttcagcgaaa	28500
ttttcgggaa	aaatggctat	tttgtgtttg	aaattgcgaa	atttcacgat	ttcgtcttaa	28560
atacgggtgc	aacctacccc	atgacggttt	gatctacaaa	aaacgcggga	atttttcaca	28620
caaaaatatg	tgagacgtct	gcacgttctt	aaccaatcgg	ttgaaaactc	tgcgcatttt	28680
ttgtagatct	acggtagatc	actgcagatt	ttaagagaga	aaaataaata	aataatccca	28740
caagggtttt	aaaatttttt	tttcaatcgt	aaaaaatagc	gaaaaattgt	ttttcgcgtc	28800
gagaccctac	gcacattttt	ttgcaatttt	cgcttcaaaa	ttacgggtacc	gggtctcgac	28860
acgacatttt	tattgtgtaa	aatacacaat	tttttggaa	tttcatcgat	tccaatttaa	28920
atatttttaa	atgatttaaa	taattcttaa	cgaaaaaaa	aaagttcgaa	actgcagtac	28980
tctttaaagg	cgcacacatg	tatgtattta	taaaaaatgt	cgtgtcaaga	ccgtactttt	29040
ggctcacaaa	ttgcaaaaata	ttgcgggaatt	ttttttaatt	ttagataaaa	aaaaacatga	29100
aaaatctatg	gaaactaaac	ttataattta	aaaaaaaatt	tttttaaggt	ggactacgct	29160
cagtggggaa	attgctttta	aacacgccta	tgaggcccca	atgactgaat	atcatgatta	29220
aaacaatcaa	aaaaaatttt	ctagattttt	tatgattttt	tgaaaattgg	aaaaatcaca	29280
gttttcacct	aattcttttt	gaatttccgc	caattggatt	agttcgggtg	agcgcgctta	29340
cattattttt	aattatttat	tttattttat	ctcgttattt	gactgatttt	cttcattttt	29400
tgtgtgtttt	cctcggaaaa	aggaagaaat	aaacaagaca	aatgcaaaat	gtttgttaaa	29460
aagtaattga	aaatgcgtaa	aactttgata	ttctgagttc	cgacgacaac	aagcctgaaa	29520
ttagtataat	tcacagtttt	tctcattttc	aattactttt	taacaaacat	tttgcatttg	29580
tcttgtgtat	ttcttccatt	ttccgaggaa	aaaacataga	aatgaagaa	aatcgggtcaa	29640
ataacgagaa	taaataaaat	taatttttaa	aaagatgcaa	gtgcgctcca	ccgaacaaat	29700
ccaattggcg	gaaattcaaa	tatggaatta	ggggaaaact	gtgatttttc	ccattttcaa	29760
aaaatccatat	aaaatttgga	aaattttttt	gaattttttt	aatcatgata	ttcgggtcatt	29820
ggcgcccat	aggcgtgttt	taaagcaatt	tccccactga	gcgtagtcca	catttaattt	29880
tccaaaaacag	cacatgctaa	tcctccaagt	tattccagac	gaggcagtta	caccggcggt	29940
ggtggtcagc	aacgaatcaa	cgtgatgggt	caaccacaac	aatgcgcag	caacaatggc	30000
ggtggagtcg	gtggccaagg	aggcctccag	ggtggtccag	gaggtccgca	aggaattcgt	30060
cggccactcg	tcggacggcc	actacaacga	ggagtcgata	atcaggcgcc	gacgggttgc	30120
caggtcgttg	ttgctccgcc	gcaaggaatg	cagcaggcat	cacaaggacc	accggtactt	30180
catatgcaga	gagcggtttc	catgcaaatg	ccgacgagtc	atcatcatca	aggccaacag	30240
caggctcctc	cgcagagctc	acagcaggct	tcgcaacagg	ctccacatc	ggattctggg	30300
acgagtgtct	cgccacgaca	agcaccacca	ccacaaaact	agaattttcc	cctattatcc	30360
tattttacc	cccaaaaactc	tattaattaa	ataattttct	tcctattttt	ttcttcgtgt	30420
gaagattatt	tgtcccccac	ccaagggtgt	cggtttttcg	atttttcgac	gtttttcaaa	30480
aaaatttcga	tttttcgaaa	aattagcttc	atattttggc	tattactctg	cttttttagaa	30540
gaaatttgta	tgttttttct	tgaaaatata	agcaaaatta	gatttaaaaa	aaatcatatt	30600
ttatggttaa	ttttctgaac	atattttttca	attttcgatt	ttcacagaaa	aacatcgaag	30660
aatcgacaaa	atcgaaaaat	atgttccgaa	aattaaccat	aaaatatgat	tttttttaaa	30720
atctaattgt	gatttatattt	ataagaaaaa	acatacaaat	ttcttctaaa	aagcagagta	30780
atagccaaaa	tatgaagcta	atttttgaaa	aaacgaaaaa	ttttcgattt	tccaaagaat	30840
cgaaaaatcg	aaaaatgaca	cccttgcccc	caactatctc	tgtatattat	tcattctatta	30900
ttgattgttt	ctttttgttc	ctcgaaattt	tttgaaatta	aagtctctct	ccccaccccg	30960
atttcgttg	ctttattaat	cgcgattgat	taattgtttt	tccataaatc	cccaactatt	31020
tatctctgta	tattattcat	ttatattatt	tatcttttat	ctgtgtcgat	ttacgggtatc	31080
tccggggcgt	atgattttga	attctcttct	caaataaaat	tgtttttcat	ctaacatttg	31140
atacgtgttt	ttctgatttt	tttgtatata	tattttccat	gtatatattt	ctttttcttt	31200
tttctttgct	ccaactttat	tttaaaataat	gcttttttat	caagagattt	tttaaaaaat	31260
cgattttttt	taaagccagg	aattctgaag	aatcgaaaaa	aatggaacta	tttttcaaat	31320
aatgagaaag	tttttttttt	tcaagaaaaa	aataataaaa	ttctgatttt	tttaataaaa	31380
atttaataag	tttttgaa	ttttcattga	aaacatctaa	actattcgat	ttttgatttt	31440
aaattttgaa	aatagaattt	tttaatatat	ttttttcaaa	tcgttaaaaa	gagaatgccg	31500
gaattttttt	aaaattcttt	aaatttagaa	ataatcgga	aattttcgat	tataaaacgc	31560
tgtataaaac	gaaaaaaagt	ggattttgat	gaaagaaaaa	attttcttgt	agtttttttc	31620
agaaaaaaat	tactttttat	tctccatttt	ttgtgttgga	atttttgaga	aaaaactcat	31680
tttgaaaaaa	tcgaattttt	tatatatttt	ctaactcgtaa	aaaaaatttt	aaaaatgaat	31740

tccggttaatt	ttttaaaaaa	taatattaat	ctatagtttt	gtagttaaaa	aaatgtttca	31800
cataaaaaatc	taaaaatttt	tgatttttaa	ttaaaaaaa	atcgaatttt	ttaaaatttt	31860
tttcaaatcg	taaaaaaga	aacaataaac	aaaagaattc	cggaaaaaaa	ttatattatg	31920
attataaatt	tatagttctt	tactttttta	aaagattttta	ttttaaaaa	tctaaaaatga	31980
tcgatttttg	gtttttttaa	ataatcaaaa	atgttttgatt	tttttttaac	gtgaaaaaaa	32040
tgcaaagaaa	atgaaatccg	gcaaaaattg	taatataatt	ataaatctat	acttttgtgg	32100
ttttttccaa	tatttctata	aattcttgat	ttttaaaata	atcaaaagtt	ttgatttttaa	32160
attttggaag	aattgaattt	ttgtatat	ttctaatacgt	aaaaaaattt	ttaaaaaaat	32220
cgaagcgga	tttttttctg	ctattttgtt	ttttttttga	aaaccggaaa	aaataccaaa	32280
aattgatagt	ttcgaccact	ctggctagac	taccaaattt	gaattttttt	tttcgaattg	32340
agaatggccg	tggtctcatc	agtagctagc	cattctcttt	ttatttcaat	ttttaagaaa	32400
aaagtctcta	aaatttttga	aaaatcgatt	ttttttactt	actttgatac	tttttttata	32460
tcttttcaaa	tcttaaaaaa	caatttttaa	aattgaattc	cggaaatttt	tttaaatat	32520
ataaatctat	agtttttttag	tttttaaaaa	atatattttt	ataaaaaatct	aaaaagtctg	32580
gcttttgact	tttgaaataa	tcgaaaatgt	ttgtttttaa	ttttgaaaaa	atataaaaaa	32640
ttcgattttt	tcaagataaa	aaagcgaatt	ttttgaattt	ttttcaaatac	gtaaaaaatg	32700
tctgtagttt	ttttaaaagac	tctcataaaa	atctgaaatg	ttcgattttt	tattttttaa	32760
ataattttta	aaaaattttt	atattttttt	tcgtgcgaat	tttttaccac	ctataatttg	32820
gaataatttt	caggatctca	aaatatccca	caatcgcgca	aatatgccag	gaagcaatga	32880
agattggata	aagaaggagg	tcgaggacca	ggacaccaac	gccaacagct	cgagctccag	32940
catagccgtc	tcgcgtcagc	tcgaaggga	ttctgctgtt	cctgacgcca	tcgaccttct	33000
gtcttctcaa	atcaaaaagag	aagttgaaga	ggaggatgat	cgcaacgatg	agactggacc	33060
ccgttcggag	cccggtggatg	ttaagccgtc	tccaaaacgc	ccaacgaaga	ggtcagccga	33120
gacctggacg	acggctcggc	gccaaagcaag	aaacggctca	cggcgaggaga	gggttcaact	33180
catcgattcg	cgtatgtgaa	tggtggagtc	cgccatccat	acgatccacg	ccatcttgtc	33240
atggaaactt	cattgaatga	aattaggtta	ggaattattg	aaaataatta	ttatatatcc	33300
atttttaattc	aatttttttt	ttcagaatcg	aagatttcga	aataatccag	tatcttccga	33360
tgcccttcag	gacttcgatt	cccatgaagc	tagtgatctt	cgcagtgaga	agtgaagaat	33420
ctgccgagaa	gatccgctcg	ttaatcgatc	cttcgatgtg	gatcgcggt	tttggtggcg	33480
gaaccgaaac	tcaaaaattc	ttgtggagcg	agctgacggg	ggaggatttc	gtcaaggcac	33540
acataatggc	cagcaggtaa	gctttcgaac	atacttaatt	ttttaaaaa	taaaattcag	33600
cgcaaccgat	gacgtgccat	atgaggcagc	catggcggat	cgagaatcgc	tcaaaacaagc	33660
tgtaaatgat	gccagctctc	tgaaaggctt	gaaggaggta	ataattttaga	aatgacagaa	33720
aatgaaccgt	gatgacgaaa	tacatctgta	aaaaaattat	aaaaaattct	aagctccgtt	33780
tttaattttt	tttttcagtt	atattctgtc	atagcggcct	atttctctgg	aaaaaaaaat	33840
ccaaaatagc	ctcaaatcgc	gaattatgct	tcgatttttt	ttctgcggta	gtcctgaatt	33900
taagacgatt	ttgaattttt	gtagctgcct	ttcgccacaa	ttacgttaaa	catttcagag	33960
catgtcgaaa	gctggatgga	ggatcgtgag	taagatgcgg	aaagatctca	atggagcctg	34020
atgatccctt	tcccagcaca	caagacagtt	ttaattttgt	gtctgtatag	ttttatatta	34080
agttttgatg	ataatgaatt	tttttacggt	tttatccatc	acttggtctg	attgaagctc	34140
ctattgtgca	gcacacacgg	cgtgtaaatt	agtgcatcta	acctaggaaa	ttcgatttct	34200
aggccatggc	cgaggatccg	actagatctt	ttttgatggt	gtttgtacag	agttaaattt	34260
cattttggag	ggaaattgaa	ggaaattgaa	agagaaatta	atttaataat	attaatttga	34320
tttaaatgac	cagaacaaaa	caaataaact	gaatgacaag	ccaatcgata	ttcgtccaga	34380
ctgggatgat	gttatatgaa	ctctttcacc	tgaaacattt	aagttttttt	aataaaaagag	34440
caagcgcgct	caaacgcgaa	aacgctcgat	ccacttaatc	tggattttgt	gccgattcat	34500
ttatttcaag	ctatgctcgt	ttttttctgt	tatgtttcat	taaaaagacc	gaaaacataa	34560
caaaaagtgc	ctgaaaacga	aaaaaaaccg	gcgacattaa	ttgaaaaatt	caaaaactaca	34620
atttcgccgc	caaaaaccaa	cgagacccaa	agtttcagcg	cggagcggtt	ccacttggcc	34680
gtggagcgcg	cttgatatata	aaaggactta	attttttaaa	atacttaccg	cagttacttc	34740
caatgtatgt	caaattcact	cgattctcca	ttgcagggtt	actaaaatat	gctccaaata	34800
gttggaagg	cgttgacttg	aatgaaatcg	gatgggttatc	ttggatgatt	gcagttcgat	34860
ttccttttgt	aattatgttc	taaaaagtca	ttgtaatcat	ttaaaagtgg	agtagcgcca	34920
gtggggattt	tgtctaaatg	cacttattat	gatccaaaac	aaccgaatat	catcataaaa	34980
cactccaaaa	agtttagttt	tttcataatt	tcctgtcaaa	gttttggaag	attggcaaaa	35040
ttttgaaaaa	tgcgagcttt	tgaggtaatt	taaggaaatg	tcgcatgttt	cgacccctac	35100
aattattttta	tacagataat	ttaaacaaaa	ttaaaacata	aaaatgtaga	aatttttttt	35160
gttttggtcg	atttccaaaa	ttatgagtgg	caaaaactga	gtaattgcc	cttttttgaca	35220

```

gtaaataaaa aatgttcaaa attttttgaa acgttttata atgatatttg gccattatgg 35280
gagcaaatga gtggtttatc tattttttca ctggcgctac tccaccttta agcatgtctg 35340
cctcaccata atcccattta atccaacggt tcttagattt ggattcgaat atatttgaat 35400
gactggaaaa tatgttacgt taccattcaa tgcaccaata taagtcatth gatcgagaaa 35460
attcaaatcg gtgagatttg tgtttctgat agtcaatggt ccgaataaaa attgtaacac 35520
tcttaatttg gaaacatatt tttcatcttc atggtctatt aatagatctc caaggatata 35580
catacatgta tctgatagtt tgctcattga ttcaaatgtg caataaaatg acgcatccaa 35640
tggaaccagga tctttgcaaa gtttcgcttc aatgttttca gtagaaattc caaggttcaa 35700
tagggcaact atctcagtaa tggtgacaca aaaatcagga tgaagggttt caaaaattgaa 35760
gtattgcctt ttattgtatg tactgtattg tatcatactg gtttgctcaa ctgtatctat 35820
aactttctga aattttatgt cattatthtc agaaatcgca ctaggcaggc aagcctgcct 35880
taccgtcaga attggcagtc ccagtcgaat catttccgga ttatcttgta cattcaatgc 35940
tacactagct atatccgagt tatattcgat agtttcgagg ttttgtaaaa acgacaaact 36000
ctgtagatta gtgttccgaa ttgcaataga tcttcgaatc attgtgacat tcaaaaatga 36060
atcataatcg aagggtgcat taatattcac taaatttaga ccagaatcta gaggtttgca 36120
tttgaggtac tcttaacat ttgatacatt aactttttca ccatcacatc ctgaaatttg 36180
actatthtta tactgttaaa aaattgtttc tcaccacaat cctttaagtt ccctctgaca 36240
atgagctcat tatacatgtg taaaaagccg ccatcacagg aaaattccag tttcggatta 36300
ttctcgattc taatatcaca cgcctcgata ccccgatcac ggtacaagta gagatcgtag 36360
agcacactgg ggtcgtttaa ttgtgaattg tttcggatgt aaacaccgtc tgaaatctga 36420
agtttaagaa aaaattaagt aagttttaat ctacatgttg atccgttttt gttgaaagta 36480
tcaaaaaact aactggagtc agaattgtctc atttcgtttt gatcttcaaa aaatgcggga 36540
gttcagacct agacatctcg tctgatttcg catggttaag agcgttctga agcgttcaat 36600
tttctgaaaa aatattcccg cattttttgt agatcaaat aaaaatgagac agcctgacac 36660
cacgtggagt tcttatata caaaaaagtt gatttttcgc tctgtatttt tctgtgtaac 36720
atcatgaaaa atccagtgtt ctctgcaaac cactaaaatc cacttttttg tttcagccgc 36780
tccgcaagca gcttcgtcga ggtcatggca gcggccgagt ttcccactcc gctgaaactc 36840
ggcacttaat atatgaacga ctaagctagc agggccgcca ttctacccta ccagcaaaaa 36900
tgaattcggt cacttacaca catcacacac cacattaaag tttccttttt ctttgtcagc 36960
tgtaaaaaacc gaaaggcttg tcagactagt atttcaata ttaaatc 37007

```

<210> 22

<211> 5656

<212> DNA

<213> *Caenorhabditis elegans*

<400> 22

```

atgccggcaa caccggtgcg tgcttcaagt actcgaataa gcagacgtac atcatcaaga 60
tcagtggctg atgatcagcc atcaacttcg tctgcggtgg ctccacctcc ttcaccatt 120
gccatagaaa ctgatgaaga tgcggtagtt gaggaggaga aaaagaagaa aaagacatca 180
gatgatttgg aaattatcac tccaagaact ccagtcgata ggcaattcc ctacatttgc 240
tcgattcttt tgactgaaaa tccgatcgatt cgcgataaat tggttctgag cagcgtcca 300
gttcgtcaag aagatcacga agaacagatt gctcgagctc aacggataca gccagttgtc 360
gatcaaatc aacgagtcga gcaaatcata ctcaatgggt cagtggaga tattctgaaa 420
gatcctcgat tccgagtaat ggcagatctc acaaaaagaac caccaccaac acctgcacct 480
cctcctccaa tccagaagac aatgcaaccg attgaggtga aaattgagga ttcagagggc 540
tcaaatacgg ctcaaccgag tgttctgccc agttgtggag gaggagagac gaatgtggaa 600
agagccgcca aaagagaagc gcatgtattg gtcgaatcg ccgagctccg taagaacggc 660
ttatggtcga acagtcgtct gccaaagtgc gtcgaacctg aacgtaataa aacgcattgg 720
gattatctac tggaaagaggt caaatggatg gcagttgatt tccgaaccga gacgaatacg 780
aagcgaaaaa tgcgcaagt tatagctcac gccattgcca aacagcaccg cgacaagcag 840
atcgagattg agagagccgc cgaacgggag atcaaggaga agcgaaaaat gtgtgcagga 900
atcgcaaga tggtagggga tttctggtcg tctacggata aagttgtgga tattcgagcg 960
aaggaagttc tggagtcgag gctcaggaag gcgagaaata agcatttgat gtttgaatt 1020
ggacaagtcg atgaaatgag caatattgtg caagaaggac ttgtttcatc gtcgaaatcc 1080
ccatcaattg catcggatcg agatgataaa gatgaagaat tcaaagcacc tggctctgat 1140

```

tcagaatctg	acgatgagca	gacaattgca	aacgcggaaa	agtcacagaa	aaaggaagat	1200
gttcgacagg	aagttgatgc	tcttcaaaac	gaggcaactg	tggatatgga	tgactttttg	1260
tacactttac	cgccggaata	tctgaaggct	tatggtctga	cgcaggagga	tttggaggag	1320
atgaagcgcg	agaaattgga	ggagcagaag	gctcgggaag	aagcttgtgg	tgataatgag	1380
gagaaaatgg	agattgatga	aagcccatca	tcagatgctc	aaaagccttc	cacctcaagc	1440
tcagatctca	ccgccgagca	gcttcaagat	ccaacagctg	aagacggcaa	cggatgagg	1500
catggtgtac	ttgaaaacgt	ggattacgtg	aagctcaaca	gtcaggatag	tgatgaacga	1560
caacaagagt	tggcgaatat	cgcagaagaa	gcgctgaaat	tccagccaaa	aggatataca	1620
cttgagacga	cacaagtcaa	gacgcccgta	ccattcctga	ttcgaggaca	actgagagaa	1680
tatcaaattgg	ttggattgga	ttggatgggt	acactttatg	agaagaattt	gaatggaatt	1740
cttgccgacg	agatgggcct	gggaaagacg	attcaaacga	tttccctgct	ggctcatatg	1800
gctttagtgt	aatcgatttt	gggaccacac	ttgattgttg	tgccgacgtc	tgtcattctg	1860
aattgggaga	tggagttcaa	gaaatggtgt	ccggtctctga	agatttttgac	gtatttttgg	1920
acggcggaag	agcgtgccga	gaagcgggaag	ggatggatga	agccgaattg	tttccatgtg	1980
tgcatcacat	catacaagac	ggttactcaa	gatattagag	cttttaagca	gagggcctgg	2040
cagtaccta	ttctcgatga	agctcaaaat	atcaaaaact	ggaagtccca	acgttggcag	2100
gctcttctga	atgtccgtgc	tcgacgtcgc	cttctcctga	ccggaactcc	acttcagaac	2160
tctctaattgg	aactgtggtc	gttgatgcat	tttttgatgc	caacaatatt	ctcaagtcat	2220
gatgatttca	aggattgggt	ctcgaatccg	ttgacaggga	tgatggaagg	aaatatggaa	2280
ttcaatgctc	cactaatcgg	acgacttcac	aaagtgtctc	gtccgtttat	tctgcggcgg	2340
ctcaagaagg	aagttgagaa	gcagctgcc	gagaagactg	agcatattgt	gaattgttcg	2400
ttgtcaaagc	ggcagagata	cctgtacgat	gactttatga	gtcgtagatc	aacaaaggag	2460
aatctaaagt	ctggaaatat	gatgtcgggt	ctcaacattg	tgatgcaact	ccgaaaatgt	2520
tgtaatcatc	cgaatctctt	cgagccgcgg	ccagttgttg	ctccgttcgt	cgttgagaag	2580
cttcagctgc	atgttccggc	tcgtctcttt	gaaatttcgc	agcaagatcc	ctctctctcc	2640
tcagctagtc	aaatttccga	aaatttcaat	ttatccaaaa	tcggctatca	atcttccgtt	2700
cgatctgcaa	aaccactcat	cgaagagctt	gaagcaatga	gcacttatcc	ggagccacga	2760
gcaccagaag	ttggcggatt	tcggttcaat	cggacggctt	ttgttgcaaa	gaatccgcac	2820
acggaagagt	cggaggacga	aggtgttatg	agaagtcgtg	ttctgccaaa	accaattaat	2880
ggaacagctc	aaccacttca	aaatggaaat	tcaataccac	aaaatgctcc	aaatcgcca	2940
caaacttcat	gcattcgttc	aaaaaccgtc	gtaaatacac	ttccactgac	catctccacc	3000
gatcgaagtg	gttttcattt	taatatggcc	aatgttggaa	gaggtgttgt	tcgtttggat	3060
gattcagcac	gtatgagccc	accgctcaaa	cgtcagaagc	tcaccggaac	tgcaacgaat	3120
tggagtgtat	atgttccgcg	acacgttggt	gaaaagatgg	aagaatcgag	aaaaaaccag	3180
ctggaaattg	ttcgaaggcg	atttgagatg	attcgtgctc	cgattattcc	actggaaatg	3240
gttgcgctgg	ttcgagagga	aattattgca	gaatttccac	gtttggctgt	ggaagaggac	3300
gaggttgtgc	aggagaggct	tttggagtat	tgcgagttgt	tggtgcaaag	attcggaaatg	3360
tacgtcgaac	cagtgtgac	cgatgcttgg	cagtgtcgtc	catcatcgtc	tggtcttcca	3420
tcatatattc	gcaacaattt	atcaaatac	gagctgaatt	ctcgttctct	tctctcaac	3480
acctccacta	atttcgatac	ccgaatgtcg	atctcacgtg	ctcttcaatt	cccagaactc	3540
cgtctgatcg	agtacgattg	tggaaagctt	cagacgttgg	ctgttctgct	tcgtcagttg	3600
tacctgtaca	agcacagatg	tctgatcttc	acgcaaattg	caaagatgct	cgacgttctg	3660
cagaccttcc	tttctcatca	cggttatcag	tatttccgcc	tcgacggtac	cactgggtgc	3720
gaacaaagac	aggcgatgat	ggagcggttc	aacgcggatc	ccaagggtgt	ttgcttcatt	3780
ctgtcgacga	gatccggtgg	tggtggagtc	aatctaaccg	gtgctgacac	tgtgatcttc	3840
tacgattcgg	attggaatcc	gacgatggat	gctcaggctc	aggatagatg	tcacgtatc	3900
ggacagacga	ggaatgtctc	gatttatcga	ttgatttccg	agcgaacaat	tgaggagaat	3960
attctgagaa	aggcaacaca	gaagcggcga	cttgagagat	tggcaattga	cgaggctggc	4020
ttcacacccg	agttcttcaa	acaatctgac	agtattcggg	atctttttga	tggagagaat	4080
gtggaagtga	ctgctgtggc	agatgttgcg	acgacgatga	gcgagaaaga	aatggagggt	4140
gcgatggcaa	agtgtgaaga	tgaagctgat	gtgaatgcgg	cgaagattgc	ggtggccgag	4200
gcgaacgttg	ataatgcgga	gtttgatgag	aaatcattgc	cgccgatgag	caatttgcaa	4260
ggagatgag	aggctgatga	gaagtatatg	gagttgatac	aacagctcaa	accaatcgaa	4320
cgatatgcc	tttaactttct	tgagacacag	tacaagccag	aatttgagga	agaatgcaaa	4380
gaggcagagg	ctcttatcga	ccaaaaacgc	gaagaatggg	acaaaaatct	caacgatacc	4440
gccgtcattg	acctcgacga	ttcggatagt	ctgctgctca	acgatccttc	gacttctgcc	4500
gatttttatc	agagctcaag	tcttttagac	gagataaaat	tctacgacga	gctggacgat	4560
atcatgccaa	tctggcttcc	accatcacca	ccagattcgg	atgcggattt	cgacttgaga	4620

```

atggaagatg attgtctcga tctgatgtat gaaattgaac aaatgaacga ggctcgccta 4680
ccacaagttt gtcattgaaat gagacgtccg ttggctgaaa aacagcagaa acagaacacg 4740
ttgaatgcgt ttaatgacat tctatcggca aaagaaaagg aatcgggtgta cgatgcggtc 4800
aacaagtgcc ttcaaatgcc acaatccgaa gcgatcacag cagaatctgc agcgtctcca 4860
gcatacacgg aacactcatc attctcgatg gatgatacaa gccaggatgc gaagattgag 4920
ccaagtttga ctgaaaatca acaaccacc accaccgcca ctactactac tacagtaccc 4980
caacaacaac aacaacagca gcagcaaaaa tcgtcgaaaa agaagagaaa tgataatcga 5040
acggctcaaa atcgaacagc tgaaaatggg gtgaaacgag cgacaactcc accaccatca 5100
tggcgtgaag agccagatta tgatggagcc gaatggaata tagttgaaga ttatgcacta 5160
cttcaagcag ttcaagtcca atttgcaaat gctcathtag tcgaaaaatc ggcgaaatgag 5220
ggaatggtgt tgaactggga attcgtgtcg aatgccgtta ataagcagac aagatttttc 5280
cgctcggccc gtcaatgctc aattcgatat caaatgtttg ttcggccaaa agagctcgga 5340
cagttggtgg cttctgatcc gatttccaag aaaacgatga aagtcgacct atcgcatact 5400
gaattatctc atttgagaaa aggacgaatg actacggaga gccaatatgc tcatgattat 5460
ggaatattga ctgataagaa acatgtgaat agatttaaaa gtgttcgagt ggcggaacaa 5520
cggagacctg ttcagttttg gagaggccct aaaggtagag gaggatggct tcataatagt 5580
cactgcaact ttttctcac gagggacgag aaaaagtggg ttctaggcca tggccgaggt 5640
gccgacaagt ttcagc                                     5656

```

<210> 23

<211> 1885

<212> PRT

<213> *Caenorhabditis elegans*

<400> 23

```

Met Pro Ala Thr Pro Val Arg Ala Ser Ser Thr Arg Ile Ser Arg Arg
 1          5          10          15
Thr Ser Ser Arg Ser Val Ala Asp Asp Gln Pro Ser Thr Ser Ser Ala
          20          25          30
Val Ala Pro Pro Pro Ser Pro Ile Ala Ile Glu Thr Asp Glu Asp Ala
          35          40          45
Val Val Glu Glu Glu Lys Lys Lys Lys Lys Thr Ser Asp Asp Leu Glu
          50          55          60
Ile Ile Thr Pro Arg Thr Pro Val Asp Arg Arg Ile Pro Tyr Ile Cys
65          70          75          80
Ser Ile Leu Leu Thr Glu Asn Arg Ser Ile Arg Asp Lys Leu Val Leu
          85          90          95
Ser Ser Gly Pro Val Arg Gln Glu Asp His Glu Glu Gln Ile Ala Arg
          100          105          110
Ala Gln Arg Ile Gln Pro Val Val Asp Gln Ile Gln Arg Val Glu Gln
          115          120          125
Ile Ile Leu Asn Gly Ser Val Glu Asp Ile Leu Lys Asp Pro Arg Phe
130          135          140
Ala Val Met Ala Asp Leu Thr Lys Glu Pro Pro Thr Pro Ala Pro
145          150          155          160
Pro Pro Pro Ile Gln Lys Thr Met Gln Pro Ile Glu Val Lys Ile Glu
          165          170          175
Asp Ser Glu Gly Ser Asn Thr Ala Gln Pro Ser Val Leu Pro Ser Cys
          180          185          190
Gly Gly Gly Glu Thr Asn Val Glu Arg Ala Ala Lys Arg Glu Ala His
195          200          205
Val Leu Ala Arg Ile Ala Glu Leu Arg Lys Asn Gly Leu Trp Ser Asn
210          215          220
Ser Arg Leu Pro Lys Cys Val Glu Pro Glu Arg Asn Lys Thr His Trp
225          230          235          240
Asp Tyr Leu Leu Glu Glu Val Lys Trp Met Ala Val Asp Phe Arg Thr

```

				245					250				255				
Glu	Thr	Asn	Thr	Lys	Arg	Lys	Ile	Ala	Lys	Val	Ile	Ala	His	Ala	Ile		
			260					265					270				
Ala	Lys	Gln	His	Arg	Asp	Lys	Gln	Ile	Glu	Ile	Glu	Arg	Ala	Ala	Glu		
		275					280					285					
Arg	Glu	Ile	Lys	Glu	Lys	Arg	Lys	Met	Cys	Ala	Gly	Ile	Ala	Lys	Met		
	290					295					300						
Val	Arg	Asp	Phe	Trp	Ser	Ser	Thr	Asp	Lys	Val	Val	Asp	Ile	Arg	Ala		
305					310					315					320		
Lys	Glu	Val	Leu	Glu	Ser	Arg	Leu	Arg	Lys	Ala	Arg	Asn	Lys	His	Leu		
			325						330					335			
Met	Phe	Val	Ile	Gly	Gln	Val	Asp	Glu	Met	Ser	Asn	Ile	Val	Gln	Glu		
			340					345					350				
Gly	Leu	Val	Ser	Ser	Ser	Lys	Ser	Pro	Ser	Ile	Ala	Ser	Asp	Arg	Asp		
	355					360						365					
Asp	Lys	Asp	Glu	Glu	Phe	Lys	Ala	Pro	Gly	Ser	Asp	Ser	Glu	Ser	Asp		
	370					375					380						
Asp	Glu	Gln	Thr	Ile	Ala	Asn	Ala	Glu	Lys	Ser	Gln	Lys	Lys	Glu	Asp		
385				390						395					400		
Val	Arg	Gln	Glu	Val	Asp	Ala	Leu	Gln	Asn	Glu	Ala	Thr	Val	Asp	Met		
			405					410						415			
Asp	Asp	Phe	Leu	Tyr	Thr	Leu	Pro	Pro	Glu	Tyr	Leu	Lys	Ala	Tyr	Gly		
			420					425					430				
Leu	Thr	Gln	Glu	Asp	Leu	Glu	Glu	Met	Lys	Arg	Glu	Lys	Leu	Glu	Glu		
	435					440					445						
Gln	Lys	Ala	Arg	Lys	Glu	Ala	Cys	Gly	Asp	Asn	Glu	Glu	Lys	Met	Glu		
	450					455					460						
Ile	Asp	Glu	Ser	Pro	Ser	Ser	Asp	Ala	Gln	Lys	Pro	Ser	Thr	Ser	Ser		
465				470						475					480		
Ser	Asp	Leu	Thr	Ala	Glu	Gln	Leu	Gln	Asp	Pro	Thr	Ala	Glu	Asp	Gly		
			485					490						495			
Asn	Gly	Asp	Gly	His	Gly	Val	Leu	Glu	Asn	Val	Asp	Tyr	Val	Lys	Leu		
			500					505					510				
Asn	Ser	Gln	Asp	Ser	Asp	Glu	Arg	Gln	Gln	Glu	Leu	Ala	Asn	Ile	Ala		
	515					520						525					
Glu	Glu	Ala	Leu	Lys	Phe	Gln	Pro	Lys	Gly	Tyr	Thr	Leu	Glu	Thr	Thr		
	530					535					540						
Gln	Val	Lys	Thr	Pro	Val	Pro	Phe	Leu	Ile	Arg	Gly	Gln	Leu	Arg	Glu		
545				550						555					560		
Tyr	Gln	Met	Val	Gly	Leu	Asp	Trp	Met	Val	Thr	Leu	Tyr	Glu	Lys	Asn		
			565					570						575			
Leu	Asn	Gly	Ile	Leu	Ala	Asp	Glu	Met	Gly	Leu	Gly	Lys	Thr	Ile	Gln		
			580					585					590				
Thr	Ile	Ser	Leu	Leu	Ala	His	Met	Ala	Cys	Ser	Glu	Ser	Ile	Trp	Gly		
	595					600						605					
Pro	His	Leu	Ile	Val	Val	Pro	Thr	Ser	Val	Ile	Leu	Asn	Trp	Glu	Met		
	610					615						620					
Glu	Phe	Lys	Lys	Trp	Cys	Pro	Ala	Leu	Lys	Ile	Leu	Thr	Tyr	Phe	Gly		
625				630						635					640		
Thr	Ala	Lys	Glu	Arg	Ala	Glu	Lys	Arg	Lys	Gly	Trp	Met	Lys	Pro	Asn		
			645					650						655			
Cys	Phe	His	Val	Cys	Ile	Thr	Ser	Tyr	Lys	Thr	Val	Thr	Gln	Asp	Ile		
			660					665					670				
Arg	Ala	Phe	Lys	Gln	Arg	Ala	Trp	Gln	Tyr	Leu	Ile	Leu	Asp	Glu	Ala		
	675					680						685					
Gln	Asn	Ile	Lys	Asn	Trp	Lys	Ser	Gln	Arg	Trp	Gln	Ala	Leu	Leu	Asn		
	690					695					700						
Val	Arg	Ala	Arg	Arg	Arg	Leu	Leu	Leu	Thr	Gly	Thr	Pro	Leu	Gln	Asn		

705		710		715		720
Ser Leu Met Glu Leu Trp Ser Leu Met His Phe Leu Met Pro Thr Ile						
	725		730			735
Phe Ser Ser His Asp Asp Phe Lys Asp Trp Phe Ser Asn Pro Leu Thr						
	740		745			750
Gly Met Met Glu Gly Asn Met Glu Phe Asn Ala Pro Leu Ile Gly Arg						
	755		760			765
Leu His Lys Val Leu Arg Pro Phe Ile Leu Arg Arg Leu Lys Lys Glu						
	770		775			780
Val Glu Lys Gln Leu Pro Glu Lys Thr Glu His Ile Val Asn Cys Ser						
	785		790			795
Leu Ser Lys Arg Gln Arg Tyr Leu Tyr Asp Asp Phe Met Ser Arg Arg						
		805		810		815
Ser Thr Lys Glu Asn Leu Lys Ser Gly Asn Met Met Ser Val Leu Asn						
	820		825			830
Ile Val Met Gln Leu Arg Lys Cys Cys Asn His Pro Asn Leu Phe Glu						
	835		840			845
Pro Arg Pro Val Val Ala Pro Phe Val Val Glu Lys Leu Gln Leu Asp						
	850		855			860
Val Pro Ala Arg Leu Phe Glu Ile Ser Gln Gln Asp Pro Ser Ser Ser						
	865		870			875
Ser Ala Ser Gln Ile Pro Glu Ile Phe Asn Leu Ser Lys Ile Gly Tyr						
		885		890		895
Gln Ser Ser Val Arg Ser Ala Lys Pro Leu Ile Glu Glu Leu Glu Ala						
	900		905			910
Met Ser Thr Tyr Pro Glu Pro Arg Ala Pro Glu Val Gly Gly Phe Arg						
	915		920			925
Phe Asn Arg Thr Ala Phe Val Ala Lys Asn Pro His Thr Glu Glu Ser						
	930		935			940
Glu Asp Glu Gly Val Met Arg Ser Arg Val Leu Pro Lys Pro Ile Asn						
	945		950			955
Gly Thr Ala Gln Pro Leu Gln Asn Gly Asn Ser Ile Pro Gln Asn Ala						
		965		970		975
Pro Asn Arg Pro Gln Thr Ser Cys Ile Arg Ser Lys Thr Val Val Asn						
	980		985			990
Thr Val Pro Leu Thr Ile Ser Thr Asp Arg Ser Gly Phe His Phe Asn						
	995		1000			1005
Met Ala Asn Val Gly Arg Gly Val Val Arg Leu Asp Asp Ser Ala Arg						
	1010		1015			1020
Met Ser Pro Pro Leu Lys Arg Gln Lys Leu Thr Gly Thr Ala Thr Asn						
	1025		1030			1035
Trp Ser Asp Tyr Val Pro Arg His Val Val Glu Lys Met Glu Glu Ser						
		1045		1050		1055
Arg Lys Asn Gln Leu Glu Ile Val Arg Arg Arg Phe Glu Met Ile Arg						
	1060		1065			1070
Ala Pro Ile Ile Pro Leu Glu Met Val Ala Leu Val Arg Glu Glu Ile						
	1075		1080			1085
Ile Ala Glu Phe Pro Arg Leu Ala Val Glu Glu Asp Glu Val Val Gln						
	1090		1095			1100
Glu Arg Leu Leu Glu Tyr Cys Glu Leu Leu Val Gln Arg Phe Gly Met						
	1105		1110			1115
Tyr Val Glu Pro Val Leu Thr Asp Ala Trp Gln Cys Arg Pro Ser Ser						
		1125		1130		1135
Ser Gly Leu Pro Ser Tyr Ile Arg Asn Asn Leu Ser Asn Ile Glu Leu						
	1140		1145			1150
Asn Ser Arg Ser Leu Leu Leu Asn Thr Ser Thr Asn Phe Asp Thr Arg						
	1155		1160			1165
Met Ser Ile Ser Arg Ala Leu Gln Phe Pro Glu Leu Arg Leu Ile Glu						

1170	1175	1180
Tyr Asp Cys Gly Lys	Leu Gln Thr Leu Ala Val	Leu Leu Arg Gln Leu
1185	1190	1195
Tyr Leu Tyr Lys His	Arg Cys Leu Ile Phe	Thr Gln Met Ser Lys Met
1205	1210	1215
Leu Asp Val Leu Gln Thr	Phe Leu Ser His His	Gly Tyr Gln Tyr Phe
1220	1225	1230
Arg Leu Asp Gly Thr Thr	Gly Val Glu Gln Arg	Gln Ala Met Met Glu
1235	1240	1245
Arg Phe Asn Ala Asp Pro	Lys Val Phe Cys Phe	Ile Leu Ser Thr Arg
1250	1255	1260
Ser Gly Gly Val Gly Val	Asn Leu Thr Gly Ala	Asp Thr Val Ile Phe
1265	1270	1275
Tyr Asp Ser Asp Trp Asn	Pro Thr Met Asp Ala	Gln Ala Gln Asp Arg
1285	1290	1295
Cys His Arg Ile Gly Gln	Thr Arg Asn Val Ser	Ile Tyr Arg Leu Ile
1300	1305	1310
Ser Glu Arg Thr Ile Glu	Glu Asn Ile Leu Arg	Lys Ala Thr Gln Lys
1315	1320	1325
Arg Arg Leu Gly Glu Leu	Ala Ile Asp Glu Ala	Gly Phe Thr Pro Glu
1330	1335	1340
Phe Phe Lys Gln Ser Asp	Ser Ile Arg Asp Leu	Phe Asp Gly Glu Asn
1345	1350	1355
Val Glu Val Thr Ala Val	Ala Asp Val Ala Thr	Thr Met Ser Glu Lys
1365	1370	1375
Glu Met Glu Val Ala Met	Ala Lys Cys Glu Asp	Glu Ala Asp Val Asn
1380	1385	1390
Ala Ala Lys Ile Ala Val	Ala Glu Ala Asn Val	Asp Asn Ala Glu Phe
1395	1400	1405
Asp Glu Lys Ser Leu Pro	Pro Met Ser Asn Leu	Gln Gly Asp Glu Glu
1410	1415	1420
Ala Asp Glu Lys Tyr Met	Glu Leu Ile Gln Gln	Leu Lys Pro Ile Glu
1425	1430	1435
Arg Tyr Ala Ile Asn Phe	Leu Glu Thr Gln Tyr	Lys Pro Glu Phe Glu
1445	1450	1455
Glu Glu Cys Lys Glu Ala	Glu Ala Leu Ile Asp	Gln Lys Arg Glu Glu
1460	1465	1470
Trp Asp Lys Asn Leu Asn	Asp Thr Ala Val Ile	Asp Leu Asp Asp Ser
1475	1480	1485
Asp Ser Leu Leu Leu Asn	Asp Pro Ser Thr Ser	Ala Asp Phe Tyr Gln
1490	1495	1500
Ser Ser Ser Leu Leu Asp	Glu Ile Lys Phe Tyr	Asp Glu Leu Asp Asp
1505	1510	1515
Ile Met Pro Ile Trp Leu	Pro Pro Ser Pro Pro	Asp Ser Asp Ala Asp
1525	1530	1535
Phe Asp Leu Arg Met Glu	Asp Asp Cys Leu Asp	Leu Met Tyr Glu Ile
1540	1545	1550
Glu Gln Met Asn Glu Ala	Arg Leu Pro Gln Val	Cys His Glu Met Arg
1555	1560	1565
Arg Pro Leu Ala Glu Lys	Gln Gln Lys Gln Asn	Thr Leu Asn Ala Phe
1570	1575	1580
Asn Asp Ile Leu Ser Ala	Lys Glu Lys Glu Ser	Val Tyr Asp Ala Val
1585	1590	1595
Asn Lys Cys Leu Gln Met	Pro Gln Ser Glu Ala	Ile Thr Ala Glu Ser
1605	1610	1615
Ala Ala Ser Pro Ala Tyr	Thr Glu His Ser Ser	Phe Ser Met Asp Asp
1620	1625	1630
Thr Ser Gln Asp Ala Lys	Ile Glu Pro Ser Leu	Thr Glu Asn Gln Gln

1635	1640	1645
Pro Thr Thr Thr Ala Thr Thr Thr Thr Thr Val	Pro Gln Gln Gln Gln	
1650	1655	1660
Gln Gln Gln Gln Gln Lys Ser Ser Lys Lys Lys Arg Asn Asp Asn Arg		
1665	1670	1675
Thr Ala Gln Asn Arg Thr Ala Glu Asn Gly Val Lys Arg Ala Thr Thr		1680
1685	1690	1695
Pro Pro Pro Ser Trp Arg Glu Glu Pro Asp Tyr Asp Gly Ala Glu Trp		
1700	1705	1710
Asn Ile Val Glu Asp Tyr Ala Leu Leu Gln Ala Val Gln Val Glu Phe		
1715	1720	1725
Ala Asn Ala His Leu Val Glu Lys Ser Ala Asn Glu Gly Met Val Leu		
1730	1735	1740
Asn Trp Glu Phe Val Ser Asn Ala Val Asn Lys Gln Thr Arg Phe Phe		
1745	1750	1755
Arg Ser Ala Arg Gln Cys Ser Ile Arg Tyr Gln Met Phe Val Arg Pro		1760
1765	1770	1775
Lys Glu Leu Gly Gln Leu Val Ala Ser Asp Pro Ile Ser Lys Lys Thr		
1780	1785	1790
Met Lys Val Asp Leu Ser His Thr Glu Leu Ser His Leu Arg Lys Gly		
1795	1800	1805
Arg Met Thr Thr Glu Ser Gln Tyr Ala His Asp Tyr Gly Ile Leu Thr		
1810	1815	1820
Asp Lys Lys His Val Asn Arg Phe Lys Ser Val Arg Val Ala Ala Thr		
1825	1830	1835
Arg Arg Pro Val Gln Phe Trp Arg Gly Pro Lys Gly Arg Gly Gly Trp		1840
1845	1850	1855
Leu His Asn Ser His Cys Asn Phe Phe Leu Thr Arg Asp Glu Lys Lys		
1860	1865	1870
Trp Phe Leu Gly His Gly Arg Gly Ala Asp Lys Phe Gln		
1875	1880	1885

<210> 24
 <211> 11851
 <212> DNA
 <213> Caenorhabditis elegans

<220>
 <221> CDS
 <222> (1001)...(1035)

<221> CDS
 <222> (1920)...(2062)

<221> CDS
 <222> (2114)...(2190)

<221> CDS
 <222> (2241)...(2501)

<221> CDS
 <222> (2551)...(2903)

<221> CDS
 <222> (2955)...(3405)

<221> CDS

<222> (3497) ... (3631)

<221> CDS

<222> (4227) ... (4690)

<221> CDS

<222> (5293) ... (6058)

<221> CDS

<222> (6696) ... (7058)

<221> CDS

<222> (7609) ... (8338)

<221> CDS

<222> (8771) ... (8933)

<221> CDS

<222> (9511) ... (10306)

<221> CDS

<222> (10774) ... (10851)

<400> 24

```

gtcaatggaa ttctcgacgc ggatcttggt agagatgccg tcgagagaga tttgatcaaa 60
ttgcggtacg ctgaaacgga tgcaccagtt ttacaggtaa aatggaaata tacaaactca 120
aaagtaaaat tttatgaatt tcagatcaac aactcactat acacggcatc ctgggagcaa 180
gatctcgga caaatatggt tctgcagtca aaaggaaaag agatggaagt gatttcgtgt 240
acatcgacca tgatgactgc agaaaaagcc ctgttgacct cgtaagcac cgaaggatct 300
acactagccg ccaatgcaga gactgctccg aaatctgac tcagtcgaac tcaaccacgt 360
caacaatgat tttcaaaata taaattaaca tgaagctctg aaataaactc atataactgc 420
taaaataaaa ctgttgcttt tgaaaccaac atttgttaga caacctgcgt ctcacagtca 480
tttttcaata tattggcgcc gcgcacacac aaagaagaag aattcgctct catggcatgg 540
catgtgcagt cagcggccac cctgtgtaac cactgcgtat cgcactcttc cactgttttt 600
tgcaatcttg ctgtcacgtt catttcctcg tacaaccatc tcttctaccc ccgttgacct 660
ctccaccatc tcatctcaat tgtgtcgttg cctccctct cccaagtct ttctgcgtct 720
cttagtgctc ttcgagaaaa gaacgaggag agctgtgaga cgctagtagg aaacgcattc 780
tcaattcgat ataggacat tgagagagag cgagcgccgt ttcgacgtct tctagccttc 840
acatcatcca gacgacgttc acacgcacac acagccaacc ccacccttct gacaacgaat 900
agacgacgaa gaagagaaga agaaaaagaa gaaggtaccc atttttcatt ccctttttgc 960
ctccacactt cactattatc gattttgtga gcgagctcta atg ttt caa cgc aaa 1015
                                     Met Phe Gln Arg Lys
                                     1           5

```

```

gtg gta ttg cct aaa aag cg gtgagaattt gcttcagaca gaaattcgtt 1065
Val Val Leu Pro Lys Lys Arg

```

10

```

ttttttaaca agaaaaatcc ggtttcaatt gtcgtagaag gtcaattttt actttcaacg 1125
ctcttcattg acggaaaact cgtttttctt tcaaatttta aattacagag gcattttact 1185
caaggtttgt tttaatttaa attaaaaata aatttttaaa tagaaatatg gataatataa 1245
aatgttttct tcaaaaaatg cactcagggt caccaaaaaa tcgataatta aaaatacggg 1305
cgcaaaggag cgtcgtagc tgctaataca tggctttaaa acgaaatcta tcgatttttg 1365
tgtactacac acggacaagt gctccaccgt tattttttga acgagtgcgt tgcaattcca 1425
tcccattttg acgtttttct tttttttt atcaaatttt ttagcattta aagtaaagtc 1485
aatgataacc tgcaataaat aatgtaaaat tcattaaaaa ccgagagaaa aagtctaaag 1545
tcataaattt ttgataaaaa agtgattttc gaaactaaaa atcattcaaa ttaaagttga 1605
acctgattct tcaattttta ttatatatta aaagcttgat ccactcaaat aaaaggagtt 1665

```

ttttaattgag aaaaaaagca aatgaaaaaa tcgataatta aattggggcgc caacctagat 1725
 ttttaatatgt ttttggttaga aatttgtata ttttcatcac tctctgactt taagcattcg 1785
 tatttttaagg aagtgtgagc tttctaatat gttttttatt aaaaaaaaca tgtttttaac 1845
 aatctccctg tcatcccat cacctaatac actcaaataa tcaataatca caatactttt 1905
 attttttctt gcag a aca gaa atg gtc caa acg aga cga aag aca gct gca 1956
 Thr Glu Met Val Gln Thr Arg Arg Lys Thr Ala Ala
 15 20

gct gta cag gac ggt ggt gcc gtt aag gag aac aaa gcc aag cca cct 2004
 Ala Val Gln Asp Gly Gly Ala Val Lys Glu Asn Lys Ala Lys Pro Pro
 25 30 35 40

gcc cct caa acg cct aca aaa cga gca aaa cga ggt cgt ccc ccg aaa 2052
 Ala Pro Gln Thr Pro Thr Lys Arg Ala Lys Arg Gly Arg Pro Pro Lys
 45 50 55

att aag act g gtgagcgaat gactatacgg aagattgaaa attcacgtgg 2102
 Ile Lys Thr

aatacttgca g at gcc aat act ttg aat acg cca agc act tct tcc aac 2151
 Asp Ala Asn Thr Leu Asn Thr Pro Ser Thr Ser Ser Asn
 60 65 70

ttg gtc gat gac aaa ctt ctc att gag tct gaa tca cag gtaaattgat 2200
 Leu Val Asp Asp Lys Leu Leu Ile Glu Ser Glu Ser Gln
 75 80 85

tcttttctat tcaaaaatta atctaaacta tacattccag gac tcg att ctc aca 2255
 Asp Ser Ile Leu Thr
 90

aac gaa gcc gac tct ttt ctg gaa aaa gaa gtg gaa gaa atc gaa gat 2303
 Asn Glu Ala Asp Ser Phe Leu Glu Lys Glu Val Glu Glu Ile Glu Asp
 95 100 105

agt tca gat ata ctt ccc gat aaa att aat tct cca gaa aaa cca agt 2351
 Ser Ser Asp Ile Leu Pro Asp Lys Ile Asn Ser Pro Glu Lys Pro Ser
 110 115 120

gtt ttg gtg aag cgg aga tcg agt acg cgg tta aaa gtg aag act gat 2399
 Val Leu Val Lys Arg Arg Ser Ser Thr Arg Leu Lys Val Lys Thr Asp
 125 130 135

gaa gat gaa aaa gat gtt cct gtg aac ata gaa gta gcc gtt tta gaa 2447
 Glu Asp Glu Lys Asp Val Pro Val Asn Ile Glu Val Ala Val Leu Glu
 140 145 150

gaa aaa tca att caa atc gag cca aca tct ccc gct cac ccg gaa gat 2495
 Glu Lys Ser Ile Gln Ile Glu Pro Thr Ser Pro Ala His Pro Glu Asp
 155 160 165 170

cct cag gtgagctttt tttaaaaata tgtattaatc aaaattcctt catttccag cct 2553
 Pro Gln Pro

tcg act tct tct ctt cca ctg gta gaa cca att gaa gac att gtg gag 2601
 Ser Thr Ser Ser Leu Pro Leu Val Glu Pro Ile Glu Asp Ile Val Glu

175	180	185	
cca aat gag cca aca agc tct gcc gat cct cca gta tca aat att aag Pro Asn Glu Pro Thr Ser Ser Ala Asp Pro Pro Val Ser Asn Ile Lys 190 195 200 205			2649
gat gag gat att aaa gaa gaa gag cca ctg att aaa aag cca gct tcc Asp Glu Asp Ile Lys Glu Glu Glu Pro Leu Ile Lys Lys Pro Ala Ser 210 215 220			2697
gat gag tca gaa tct atg gat ata gct aac tct gaa agt gga aat gat Asp Glu Ser Glu Ser Met Asp Ile Ala Asn Ser Glu Ser Gly Asn Asp 225 230 235			2745
tcc gat tca agt gaa gct gat cct agg acg ata cca tct ttc tct ata Ser Asp Ser Ser Glu Ala Asp Pro Arg Thr Ile Pro Ser Phe Ser Ile 240 245 250			2793
cct ctt ccc gac aca cca cct cca aat ttt gcg aaa aga gga gaa ata Pro Leu Pro Asp Thr Pro Pro Pro Asn Phe Ala Lys Arg Gly Glu Ile 255 260 265			2841
cat gta gat gta gat cag aaa aat tcc aag caa tca gga gaa tca caa His Val Asp Val Asp Gln Lys Asn Ser Lys Gln Ser Gly Glu Ser Gln 270 275 280 285			2889
tcg cct tgg gag cg gtaagaatat ttatcctagc caggtgttat aacaaaattg Ser Pro Trp Glu Arg 290			2943
aatagtttca g a gca aga gaa aag tct gca tcg aac cca ttg tcc tct Ala Arg Glu Lys Ser Ala Ser Asn Pro Leu Ser Ser 295 300			2991
cca aca atg agc cga ccc agg ata cac ttc ctt cat cca gca tat caa Pro Thr Met Ser Arg Pro Arg Ile His Phe Leu His Pro Ala Tyr Gln 305 310 315			3039
agt ttc aca aat gat tca gtt tca cct cta cca cca ccg cca cca gag Ser Phe Thr Asn Asp Ser Val Ser Pro Leu Pro Pro Pro Pro Glu 320 325 330			3087
ccg gct cca gct cgt gaa aaa gtg gaa aat ggt ggt cca act act ttc Pro Ala Pro Ala Arg Glu Lys Val Glu Asn Gly Gly Pro Thr Thr Phe 335 340 345 350			3135
aaa atg act ttc aaa aaa gct gca aat att cct atc ttg aag aca tcg Lys Met Thr Phe Lys Lys Ala Ala Asn Ile Pro Ile Leu Lys Thr Ser 355 360 365			3183
gca ttt gaa caa cca tca tca cct cca cct tcc tca tca gtt tct tca Ala Phe Glu Gln Pro Ser Ser Pro Pro Pro Ser Ser Ser Val Ser Ser 370 375 380			3231
tca att tca tta tct gaa gtg aat tct tct aca tcg ata gcc tcc gag Ser Ile Ser Leu Ser Glu Val Asn Ser Ser Thr Ser Ile Ala Ser Glu 385 390 395			3279

```

tct tct cca gcg aaa aga agc tca aat ttc gat tta act gcc tca aat 3327
Ser Ser Pro Ala Lys Arg Ser Ser Asn Phe Asp Leu Thr Ala Ser Asn
400 405 410

gag ctt cca cca cct cag atg gtt gaa ctt ccc aag ctc tca ttt ttc 3375
Glu Leu Pro Pro Pro Gln Met Val Glu Leu Pro Lys Leu Ser Phe Phe
415 420 425 430

aat atg cct cca gcc gtt cgc tcc gca gag gtttagttaac tttttcccg 3425
Asn Met Pro Pro Ala Val Arg Ser Ala Glu
435 440

tttcatgaaa tttcagcggg atctgtcttc cttttggtgt gtgccctcac aacctaacct 3485
cttttatecca g gac gat tct gcg atg acg tcg gaa gaa ccg atc ctt ctc 3535
Asp Asp Ser Ala Met Thr Ser Glu Glu Pro Ile Leu Leu
445 450

ctc cgt tct ccg aat tcc gcc act cct gat gat gat gca ctt ttc ctc 3583
Leu Arg Ser Pro Asn Ser Ala Thr Pro Asp Asp Asp Ala Leu Phe Leu
455 460 465

acg acc cca cca cca ccc aag atg acc gaa tca gaa att caa gca ctg 3631
Thr Thr Pro Pro Pro Pro Lys Met Thr Glu Ser Glu Ile Gln Ala Leu
470 475 480 485

gtgagccaga tcacacattt cgatgtcgtg tgtggaaccc aggaatttca gaccgtttttt 3691
ctttacacct catccccctt tgtgttatgt taacattcat tttgtgtctc aaacactgca 3751
tgcttttgca cttggaaatt aaaaaataat gcgttctggg attttgtgtg ttaagggtgga 3811
gtagagtttg tgaggctaga aagtatgcct ttttcgtttc tccactgcaa aatttcgttt 3871
gaaaaaaaca aaaaatttac taaaatttga aatttcacca acttgccggt gtcacagctg 3931
ctgaaatata gtttttattg cattttcacc ctttattgca tattattatt agacaccttt 3991
taggtcaata ggcaaccgaa aatatccgaa ttgacttaa aatgtacctt aattaaggaa 4051
ctaacttgag atatacgact aaaaatgcaa taaattgtga gaattattgt tatgaaattc 4111
agccgtttta ggctagtttt agccaaaaac cgacaaactc tattccaatt aattttccac 4171
tctgcacct cgattagtga ttttttgaag aaaaaaaatt atcttcttat ttcag aaa 4229
Lys

gta gcg acg gaa aaa gtg aat caa gta att gct cga cgt gaa gat tct 4277
Val Ala Thr Glu Lys Val Asn Gln Val Ile Ala Arg Arg Glu Asp Ser
490 495 500

gaa aaa gat gta cgt cac aga gaa gat cga gat gat tat gat aga cga 4325
Glu Lys Asp Val Arg His Arg Glu Asp Arg Asp Asp Tyr Asp Arg Arg
505 510 515

cgt gac gac cgt gac aga aga tcc aga aag act gat tcg gaa cga aat 4373
Arg Asp Asp Arg Asp Arg Arg Ser Arg Lys Thr Asp Ser Glu Arg Asn
520 525 530

gat caa aga gga cga caa cgt gaa gat gat gaa cga aga gct cga gaa 4421
Asp Gln Arg Gly Arg Gln Arg Glu Asp Asp Glu Arg Arg Ala Arg Glu
535 540 545 550

cga gaa aga gaa gtt acg aaa cga cat gat cgg gaa agg gaa gag atg 4469
Arg Glu Arg Glu Val Thr Lys Arg His Asp Arg Glu Arg Glu Glu Met
555 560 565

```

cga tta cag aaa caa aaa gat gag gaa aga aga aag aaa gat gaa gag	4517
Arg Leu Gln Lys Gln Lys Asp Glu Glu Arg Arg Lys Lys Asp Glu Glu	
570 575 580	
 gaa agg ata caa aaa gag aat gat gag aaa aaa caa aaa gag gat gaa	4565
Glu Arg Ile Gln Lys Glu Asn Asp Glu Lys Lys Gln Lys Glu Asp Glu	
585 590 595	
 gcc aaa atg gag gag gag aaa aag aag att aaa gag gag gaa atg aag	4613
Ala Lys Met Glu Glu Glu Lys Lys Lys Ile Lys Glu Glu Glu Met Lys	
600 605 610	
 att cct gaa ttt gag ttg att agc gaa tca aaa tat ttg acg agg aat	4661
Ile Pro Glu Phe Glu Leu Ile Ser Glu Ser Lys Tyr Leu Thr Arg Asn	
615 620 625 630	
 gcg aat aaa aag aag act gaa tcc tta ac gtaagttatt atttataaat	4710
Ala Asn Lys Lys Lys Thr Glu Ser Leu Thr	
635 640	
 ttgacttaaa aattgataac tttcaaaaatt aagtgattca atagactcaa aagaatgaaa	4770
aactagagtg cgcttttaaa gagtactgta atttcaaact tttgttgctg ctcatttttc	4830
atcgattttt cttagttttt cgttaaaaat aattcaacca ttggattaaa aaaaattaaa	4890
aacacataaa ttttattttg aaaagtaatg agaaaaacta tagaaattcg ccgaaaattc	4950
tacagcaaca aaagctcaaa attacagtac tttttaaagg agcacatctt tctgaattta	5010
acaaaaattc ggagattttt ctttttttcg tgtttttctg gcgaaaaaac gatttttcgc	5070
ttttaccgga aacggtatcc ggaggaaaaa aaaaacgaaa aaagcgaaaa attttaagaa	5130
gtttcaagat tagttacaaa ctcttttcaa aagcagattc tacagttttt tgggggtttg	5190
ccaaaaaatt tatgaaatat aatgtttttt agactagaaa aataaactaa ttttaatttt	5250
caatcaaaag ctcattatta tatttatatt tatataattc ag t tgc gaa tgc cat	5305
Cys Glu Cys His	
 cga act ggt gga aac tgt tgc gac aat act tgt gtg aat cgt gca atg	5353
Arg Thr Gly Gly Asn Cys Ser Asp Asn Thr Cys Val Asn Arg Ala Met	
645 650 655 660	
 ctc acc gag tgc cca tca tca tgt cag gtc aaa tgc aag aat caa cga	5401
Leu Thr Glu Cys Pro Ser Ser Cys Gln Val Lys Cys Lys Asn Gln Arg	
665 670 675	
 ttt gca aag aaa aag tac gcg gct gtt gaa gca ttc cac act gga acc	5449
Phe Ala Lys Lys Lys Tyr Ala Ala Val Glu Ala Phe His Thr Gly Thr	
680 685 690	
 gcc aaa gga tgt gga ctt cga gca gtg aaa gac ata aaa aaa gga aga	5497
Ala Lys Gly Cys Gly Leu Arg Ala Val Lys Asp Ile Lys Lys Gly Arg	
695 700 705	
 ttc atc att gaa tat ata gga gaa gtt gtg gaa aga gat gat tat gag	5545
Phe Ile Ile Glu Tyr Ile Gly Glu Val Val Glu Arg Asp Asp Tyr Glu	
710 715 720	
 aag aga aaa acg aaa tat gca gct gat aaa aag cac aaa cat cat tat	5593
Lys Arg Lys Thr Lys Tyr Ala Ala Asp Lys Lys His Lys His His Tyr	
725 730 735 740	
 ctc tgt gat act gga gtc tac acg atc gac gca aca gtc tac gga aat	5641

Leu Cys Asp Thr Gly Val Tyr Thr Ile Asp Ala Thr Val Tyr Gly Asn	
745 750 755	
cca tct cga ttt gtg aat cat agt tgt gat cct aat gct ata tgt gag	5689
Pro Ser Arg Phe Val Asn His Ser Cys Asp Pro Asn Ala Ile Cys Glu	
760 765 770	
aaa tgg tct gta cca aga act cct gga gac gtt aat cga gtt ggt ttc	5737
Lys Trp Ser Val Pro Arg Thr Pro Gly Asp Val Asn Arg Val Gly Phe	
775 780 785	
ttc tcg aaa cga ttc att aaa gcc ggc gaa gaa atc aca ttt gat tat	5785
Phe Ser Lys Arg Phe Ile Lys Ala Gly Glu Glu Ile Thr Phe Asp Tyr	
790 795 800	
caa ttt gtc aac tac gga cgt gac gct caa caa tgt ttc tgt gga agt	5833
Gln Phe Val Asn Tyr Gly Arg Asp Ala Gln Gln Cys Phe Cys Gly Ser	
805 810 815 820	
gct tca tgt agt gga tgg att ggg cag aaa ccg gaa gaa ttt tca tct	5881
Ala Ser Cys Ser Gly Trp Ile Gly Gln Lys Pro Glu Glu Phe Ser Ser	
825 830 835	
gat gag gat gat gat att gtg act aca agg cat att aat atg gat gaa	5929
Asp Glu Asp Asp Asp Ile Val Thr Thr Arg His Ile Asn Met Asp Glu	
840 845 850	
gaa gaa gaa gaa aag ttg gaa ggt ctt gat cat ctt gga aat cat gaa	5977
Glu Glu Glu Glu Lys Leu Glu Gly Leu Asp His Leu Gly Asn His Glu	
855 860 865	
cgg aat gaa gtg atc aag gat atg ttg gat gat ttg gtc att cgg aat	6025
Arg Asn Glu Val Ile Lys Asp Met Leu Asp Asp Leu Val Ile Arg Asn	
870 875 880	
aag aag cat gct agg aag gtt atc aca att gcg gtaagcattt atttgttagag	6078
Lys Lys His Ala Arg Lys Val Ile Thr Ile Ala	
885 890 895	
aaaatttaaa aattaaagat ggagtaccga aatccgagaa atatatttaa ttgactccaa	6138
tttttcctct gattccgaat ttttaaataa aaaaattcaa aaaaatttcc ttgattttat	6198
gttttaactt gaaattgcga atttcatttg tacagatttt tgaaacgccg aattttcgcg	6258
ccagagaagc catgtgtcga tttttgagat ttgtgtatat ttacaagatt ttgaatcttc	6318
atcggatgct gat ttg cgt tttcatcatt atattatcaa aaaactaaca atttggtcgg	6378
ttttacggaa attaacaata tagactagac atttcgtaaa tatacacaaa tctcgtaaat	6438
cgacacatgg cgtctctggc gcgaaaattc ggcatttgaa aaatcttatg cgggcactaa	6498
tgaaattcgt gatttcaagc tgaaatataa aatcagggaa ttttccttgc attttttcac	6558
tcagaacttc ggaatcagtt gcaaatttgg agtcatttga aaatatttct cagatttcgg	6618
tactccacct ttattataat ttttaaaatt ttttaaataa ttttttttcc atgttcaaca	6678
aaaaaataaa ttttcag tct gca atg acc gat tac tct caa cgt gtg gat	6728
Ser Ala Met Thr Asp Tyr Ser Gln Arg Val Asp	
900 905	
gtc att caa gaa atc ttc tcc tca gac acc tcc gta acc gtt caa aaa	6776
Val Ile Gln Glu Ile Phe Ser Ser Asp Thr Ser Val Thr Val Gln Lys	
910 915 920	
ttc tat gca aaa gag gga atg gct aca ttg atg gct gaa tgg ttg tct	6824

Phe Tyr Ala Lys Glu Gly Met Ala Thr Leu Met Ala Glu Trp Leu Ser	
925 930 935	
gaa gat gat tat tcg ctg gat aat ctg aaa ctt gtt caa gct att ctc	6872
Glu Asp Asp Tyr Ser Leu Asp Asn Leu Lys Leu Val Gln Ala Ile Leu	
940 945 950	
aaa gct ctt cac act gaa cta ttc gat tcg tgc gcc aaa aat gat cga	6920
Lys Ala Leu His Thr Glu Leu Phe Asp Ser Cys Ala Lys Asn Asp Arg	
955 960 965 970	
ctc tta cga gat tct aca tca cga tgg gtc aat gcg aaa atg gat gaa	6968
Leu Leu Arg Asp Ser Thr Ser Arg Trp Val Asn Ala Lys Met Asp Glu	
975 980 985	
tat gtt gat ata caa gtg ata gct gat tca ctt att gct tgt gtt gaa	7016
Tyr Val Asp Ile Gln Val Ile Ala Asp Ser Leu Ile Ala Cys Val Glu	
990 995 1000	
gat ccc gta cag gag tac aag gat gtt tgc aaa gtt ata gag	7058
Asp Pro Val Gln Glu Tyr Lys Asp Val Cys Lys Val Ile Glu	
1005 1010 1015	
gtatatacat attaatTTTT aaaaaagaat attTTTTtgca tgtcacaaaa tatttgaaaa	7118
TTTTcccgaa aaacccatga aatcaaaaaa caaattaaat agtaaaatta tttcctccta	7178
cgaacatttt tCGatttttc gttttccgat attcctttta aaaatctgat ttaaaaaaaa	7238
aaaacttaaa ttttaggtct ttttgctcct ttttagaagc aatttatatg ttttttaaaa	7298
caaaacttaa aattagcatt tttatgggta attttctgaa cacatttttt tttcgaaaaa	7358
aatggccaga atttcaacca cttctccgta aaatcgaaat taactaattt tttctctata	7418
cattttttcaa aaaaagactc ctcattttatt gtattagata caaatatatg ttttcctcat	7478
caaaattttac gaaatttggtt ataattttga attttttttg tttttttttc gaaaaattga	7538
aaatttttcta attttgaaac gatattatac aatttcagcg ccatcaattt aactaattaa	7598
ataatttcag aaa ggt ctc gtc gaa aac ttc aca aga gcc aaa gag atg	7647
Lys Gly Leu Val Glu Asn Phe Thr Arg Ala Lys Glu Met	
1020 1025	
gcc tat cgg tta aat caa tac tgg ttc aat cga tca gtg agc ttc aaa	7695
Ala Tyr Arg Leu Asn Gln Tyr Trp Phe Asn Arg Ser Val Ser Phe Lys	
1030 1035 1040 1045	
att cca aaa aag ata cgt gat cct gtg cca aaa gat gtt cca gtc aga	7743
Ile Pro Lys Lys Ile Arg Asp Pro Val Pro Lys Asp Val Pro Val Arg	
1050 1055 1060	
caa gaa gat gct aca aca tca tca caa tct cat gat aat agt agt aga	7791
Gln Glu Asp Ala Thr Thr Ser Ser Gln Ser His Asp Asn Ser Ser Arg	
1065 1070 1075	
act gta tca ccg aat cat cga cat cat tca tct tca tat tca aat tca	7839
Thr Val Ser Pro Asn His Arg His His Ser Ser Ser Tyr Ser Asn Ser	
1080 1085 1090	
tgt tat caa gaa cga gaa cca tct cat ata cga ttc ttt aat aat gga	7887
Cys Tyr Gln Glu Arg Glu Pro Ser His Ile Arg Phe Phe Asn Asn Gly	
1095 1100 1105	
aat gat gtt cat caa tat cgt ttt gga ggt tat cat gga aat aac tac	7935
Asn Asp Val His Gln Tyr Arg Phe Gly Gly Tyr His Gly Asn Asn Tyr	

1110	1115	1120	1125	
aat gat aac tat ttc agt aga agg ccc aat aag gat tca tat cga gat				7983
Asn Asp Asn Tyr Phe Ser Arg Arg Pro Asn Lys Asp Ser Tyr Arg Asp				
	1130	1135	1140	
cgc cgt cga ttt aat gga cgt cgt tcg aga agt cga tca aga agt gtc				8031
Arg Arg Arg Phe Asn Gly Arg Arg Ser Arg Ser Arg Ser Arg Ser Val				
	1145	1150	1155	
tca cca cag aac tat aaa aga aga aaa ctc gat gaa cat gac aat aat				8079
Ser Pro Gln Asn Tyr Lys Arg Arg Lys Leu Asp Glu His Asp Asn Asn				
	1160	1165	1170	
cat cgt cag cgt tct cca att cgt gat cgt cac aca tct ccc ggc ggc				8127
His Arg Gln Arg Ser Pro Ile Arg Asp Arg His Thr Ser Pro Gly Gly				
	1175	1180	1185	
gaa aag act cct agc tcg aat aat tct gga gaa cga aac tat aaa aga				8175
Glu Lys Thr Pro Ser Ser Asn Asn Ser Gly Glu Arg Asn Tyr Lys Arg				
	1190	1195	1200	1205
ctg gat att cga gga gct cgt ata aaa act ata aaa gaa gat ttg gaa				8223
Leu Asp Ile Arg Gly Ala Arg Ile Lys Thr Ile Lys Glu Asp Leu Glu				
	1210	1215	1220	
gct gct gct gct gct gct gct gct gct gta cca tca gaa gtg caa				8271
Ala Ala Ala Ala Ala Ala Ala Ala Ala Val Pro Ser Glu Val Gln				
	1225	1230	1235	
gct tat cct cat gaa cat aca gct gta cat cag agt gtt tat cag atg				8319
Ala Tyr Pro His Glu His Thr Ala Val His Gln Ser Val Tyr Gln Met				
	1240	1245	1250	
cca ggt tat gag tct tat g gttggttag tttttttaaa aatatcattt				8368
Pro Gly Tyr Glu Ser Tyr				
	1255			
accagggtgc cattttttaaa aataaaaaata actcggaaaa tatgttttta aaaaatttca				8428
gaattttctct catcaacata aaacttgata aaaatcgaat ttttattatt ttctaaacat				8488
tttttcggtt tttccgaaaa tcaaaaaaaa agtttagaaa atagcaaaaa atcagtttat				8548
tagaaatcaa attttggttcg ttttgataag aaaaaacata agaaaacatg ttattttctt				8608
ctgaaaaaag aaaaaaatcg aaaaatctat ggccttttgg caaaatgttt tggaccaaaa				8668
aacaaaacaa atagcattaa aattattagt tcttttggtt tcttctaaag ttaattttct				8728
gaaagtcttg cttgtcgtat atcaaataaa aacatttttc ag ga gta tat gat cct				8784
			Gly Val Tyr Asp Pro	
			1260	
gta aat ggt gtc tac atg tat cct cat cct ggc gct ggt tac tat cca				8832
Val Asn Gly Val Tyr Met Tyr Pro His Pro Gly Ala Gly Tyr Tyr Pro				
	1265	1270	1275	1280
cct gcc tat cca caa caa ccg att atg tta aca atg gac act ctt cca				8880
Pro Ala Tyr Pro Gln Gln Pro Ile Met Leu Thr Met Asp Thr Leu Pro				
	1285	1290	1295	
ccg aat gat cgt ctt ggt gaa ctt tac gag aaa gcc agt atc gag cag				8928
Pro Asn Asp Arg Leu Gly Glu Leu Tyr Glu Lys Ala Ser Ile Glu Gln				

1300	1305	1310	
cta gc gtgagcattt tttagtttaa acctttcgga tttacctaga aaaatgttac			8983
Leu Ala			
ctttgacgca aaattacggt agcagggtctc gtcgcgaccg aaattttttca gcgaggtacg			9043
gtagcttccc atgaattttt ttgctgaact tatctttctg ataacaaata gtaactaaaa			9103
catgaaaaac tgaataaaaa ttgatatctt taccttatag gctctttaag ggcgcagaca			9163
caaaaactga ccggctaccg taattttttcg tcaaaagtca cacatttctc aactgggtgaa			9223
atccgaaaaa attgaaattt ttactactcg tccgactggt tagaaaaagat taaaaaaaaa			9283
gaaaaaaaaa atgtcgggtt ttcgaatttt cgattttcaa agaaaaaaat caatatttaa			9343
aaatcatttt cggtaatttt cctaaatttg taaaatataa tttccaataa atgttttttg			9403
ttttccggaa ttttaataaa aaatcaattt tcgcgtaaca aaaatgcgaa aaaatgacta			9463
gccactcgaa tataataaca catgaaataa aattaaaatt attacag t caa cga gat			9520
		Gln Arg Asp	
		1315	
gca att gtg aga caa gaa ctt gag ctg ata cgt att caa atc gaa aga			9568
Ala Ile Val Arg Gln Glu Leu Glu Leu Ile Arg Ile Gln Ile Glu Arg			
1320	1325	1330	
aaa act gct caa aaa gaa gcg atc aag gcc gct tgc cgt cgt gct aac			9616
Lys Thr Ala Gln Lys Glu Ala Ile Lys Ala Ala Cys Arg Arg Ala Asn			
1335	1340	1345	
gaa gaa gaa gct aaa cga caa gag gca ctt gca aag acg aaa tat gtt			9664
Glu Glu Glu Ala Lys Arg Gln Glu Ala Leu Ala Lys Thr Lys Tyr Val			
1350	1355	1360	1365
tgg gcg att gca aag tca gaa gct gga gag acg tat tac tac aac aaa			9712
Trp Ala Ile Ala Lys Ser Glu Ala Gly Glu Thr Tyr Tyr Tyr Asn Lys			
1370	1375	1380	
ata aca aaa gag acg cag tgg aca gca cca aca cca gtt caa ggt ctt			9760
Ile Thr Lys Glu Thr Gln Trp Thr Ala Pro Thr Pro Val Gln Gly Leu			
1385	1390	1395	
ctc gaa ccg gct tgt ggt gca tct cct gat act aca gtt gtc att gct			9808
Leu Glu Pro Ala Cys Gly Ala Ser Pro Asp Thr Thr Val Val Ile Ala			
1400	1405	1410	
gac gag att act gaa gaa gag caa caa gct gaa gtt ctg gag aag ccg			9856
Asp Glu Ile Thr Glu Glu Glu Gln Gln Ala Glu Val Leu Glu Lys Pro			
1415	1420	1425	
cgt gtt gtt aag gaa gaa gtt atc gag cca ggt tca caa tct gaa act			9904
Arg Val Val Lys Glu Glu Val Ile Glu Pro Gly Ser Gln Ser Glu Thr			
1430	1435	1440	1445
caa aaa gaa tct ccg gag aaa gtt cga gtt gtt gta ccg aaa gtt gaa			9952
Gln Lys Glu Ser Pro Glu Lys Val Arg Val Val Val Pro Lys Val Glu			
1450	1455	1460	
ggt gaa aga tca ccg tcg cca aaa tct tct cgt gat cgt gag aag gat			10000
Val Glu Arg Ser Pro Ser Pro Lys Ser Ser Arg Asp Arg Glu Lys Asp			
1465	1470	1475	

cga gag aaa tct cgt gag aaa gat cgt gaa aga gat cgt gac aga aga 10048
 Arg Glu Lys Ser Arg Glu Lys Asp Arg Glu Arg Asp Arg Asp Arg Arg
 1480 1485 1490

gaa ggt tca aaa cat cgt gat agt tat cat gga cat cga aac ggc agc 10096
 Glu Gly Ser Lys His Arg Asp Ser Tyr His Gly His Arg Asn Gly Ser
 1495 1500 1505

agt tct gtc agt gaa cga cgt atg cga gag ttc aaa cat gag ctg gaa 10144
 Ser Ser Val Ser Glu Arg Arg Met Arg Glu Phe Lys His Glu Leu Glu
 1510 1515 1520 1525

cga tcc act cga tct gcc gtt cgt tct cgt cta caa cat caa cgt gac 10192
 Arg Ser Thr Arg Ser Ala Val Arg Ser Arg Leu Gln His Gln Arg Asp
 1530 1535 1540

gct tct agt gat aag act act tgg ctt att aag tta ata tat cga gag 10240
 Ala Ser Ser Asp Lys Thr Thr Trp Leu Ile Lys Leu Ile Tyr Arg Glu
 1545 1550 1555

att ttc aaa cga gaa agt gcg cag agt gga ttt gat tat cga ttc agt 10288
 Ile Phe Lys Arg Glu Ser Ala Gln Ser Gly Phe Asp Tyr Arg Phe Ser
 1560 1565 1570

gag aat act gat aag aag gtaatatattat ggaccaaaaa ataaacaatt 10336
 Glu Asn Thr Asp Lys Lys
 1575

gaaaaaaaa ccaaaaaaat ctgatgcttg aattttaaaaa aaaacaatga aagagtgcaa 10396
 ttttttaggt tttttggtct ttttttttg aaaaaccaa aaataaattt ttttccaaag 10456
 taccaaactt cattttaaaa aattttattt gacataaaaa ttgataattt aaaactaatt 10516
 tgaacatttt tccgcaaaaa ttatagattt ttctgccaat tttagatttt taacgttttt 10576
 tttcgacaa ttaatgtttc gaatcatcaa tcagaatgaa tatgatattt gatgaaattc 10636
 aaaaataatg caattttaaat agaaaacggt acaaaagtgt tgaaaaattt agaagaattc 10696
 taaaaaaaa cctgtccttc aggacaaaat tcaacctttt tctcaaaaaca caaaaattac 10756
 tttatattat ttttcag gtg aaa aac tac gtc aag tca tat atc gac cga 10806
 Val Lys Asn Tyr Val Lys Ser Tyr Ile Asp Arg
 1580 1585 1590

aaa ctc gaa tca aac gat ctc tgg aaa gaa tac tct cgg cca tga 10851
 Lys Leu Glu Ser Asn Asp Leu Trp Lys Glu Tyr Ser Arg Pro *
 1595 1600

gctttatttt ttaatttaaa ttttataaaa aaatgtttat gcttgttttt ttctctatag 10911
 ttccctccta tccccccct cccctatcgc ctaaaaattg atctctgtct gatttcaccg 10971
 atttccgttt tatttgatcc cattgaacga gtatatcatc atgttctga acttcaacgt 11031
 tcgcacattt tattcccta gttttatgtc ccagaattg ttttatacta tctgtgaatc 11091
 cacctcaaaa tgacagccat gaaaagctgt ttttcatgtt ttctatatttc ttgttgatcg 11151
 tatttgcgcc gctctttgtc gccaaatttt tttttgtaat taaaaaatga attacggatg 11211
 ttgaattttt aaattttatt ttttaaagaa aaattgtgga agtttttcag attctatact 11271
 gcttattttt acgctaaatt ttttttcgaa gtcccccttt ttcaaatoaga agtgtaactg 11331
 cgctccacga tcaatagaga ctctccgccc tcgaaccatg ggtctcgta ggtatttggc 11391
 agacttaccg taaattcaaaa tgtttttatta ctctcgact aattttttta ttcattgactc 11451
 aattttttat caattccaac gaaaaactaa ttaaaaaaca cggaaaacat aacgaaaaat 11511
 gcttgaaaat tgcagacatt tccgaaatta attaaattcc taacgagacc catggctcgg 11571
 gggcgagtg ttttcgatta gccatggagc gcgttgagat attcctaaat ttttctattc 11631
 agatgtcgaa tcaatcaaaa cgggtcacag tgagaattga gcattogaag aacacttttt 11691
 tcgaaaagta attttcaaat tttgatccaa agaaattatt cgtcaatttt cagagtttta 11751

aaattccaac atcaagagca agaagatcgg aagctcaa atgttctgca caaagctcac 11811
 gagaatctga gaaagtgcc attcgagatt ctgacaattg 11851

<210> 25

<211> 1604

<212> PRT

<213> *Caenorhabditis elegans*

<400> 25

Met	Phe	Gln	Arg	Lys	Val	Val	Leu	Pro	Lys	Lys	Arg	Thr	Glu	Met	Val
1				5					10					15	
Gln	Thr	Arg	Arg	Lys	Thr	Ala	Ala	Ala	Val	Gln	Asp	Gly	Gly	Ala	Val
			20					25					30		
Lys	Glu	Asn	Lys	Ala	Lys	Pro	Pro	Ala	Pro	Gln	Thr	Pro	Thr	Lys	Arg
		35					40					45			
Ala	Lys	Arg	Gly	Arg	Pro	Pro	Lys	Ile	Lys	Thr	Asp	Ala	Asn	Thr	Leu
	50					55					60				
Asn	Thr	Pro	Ser	Thr	Ser	Ser	Asn	Leu	Val	Asp	Asp	Lys	Leu	Leu	Ile
65					70					75				80	
Glu	Ser	Glu	Ser	Gln	Asp	Ser	Ile	Leu	Thr	Asn	Glu	Ala	Asp	Ser	Phe
				85					90				95		
Leu	Glu	Lys	Glu	Val	Glu	Glu	Ile	Glu	Asp	Ser	Ser	Asp	Ile	Leu	Pro
			100					105					110		
Asp	Lys	Ile	Asn	Ser	Pro	Glu	Lys	Pro	Ser	Val	Leu	Val	Lys	Arg	Arg
	115						120					125			
Ser	Ser	Thr	Arg	Leu	Lys	Val	Lys	Thr	Asp	Glu	Asp	Glu	Lys	Asp	Val
	130				135						140				
Pro	Val	Asn	Ile	Glu	Val	Ala	Val	Leu	Glu	Glu	Lys	Ser	Ile	Gln	Ile
145					150					155				160	
Glu	Pro	Thr	Ser	Pro	Ala	His	Pro	Glu	Asp	Pro	Gln	Pro	Ser	Thr	Ser
				165					170					175	
Ser	Leu	Pro	Leu	Val	Glu	Pro	Ile	Glu	Asp	Ile	Val	Glu	Pro	Asn	Glu
		180						185					190		
Pro	Thr	Ser	Ser	Ala	Asp	Pro	Pro	Val	Ser	Asn	Ile	Lys	Asp	Glu	Asp
		195					200					205			
Ile	Lys	Glu	Glu	Glu	Pro	Leu	Ile	Lys	Lys	Pro	Ala	Ser	Asp	Glu	Ser
	210					215					220				
Glu	Ser	Met	Asp	Ile	Ala	Asn	Ser	Glu	Ser	Gly	Asn	Asp	Ser	Asp	Ser
225					230					235				240	
Ser	Glu	Ala	Asp	Pro	Arg	Thr	Ile	Pro	Ser	Phe	Ser	Ile	Pro	Leu	Pro
			245						250					255	
Asp	Thr	Pro	Pro	Pro	Asn	Phe	Ala	Lys	Arg	Gly	Glu	Ile	His	Val	Asp
		260						265					270		
Val	Asp	Gln	Lys	Asn	Ser	Lys	Gln	Ser	Gly	Glu	Ser	Gln	Ser	Pro	Trp
		275					280					285			
Glu	Arg	Ala	Arg	Glu	Lys	Ser	Ala	Ser	Asn	Pro	Leu	Ser	Ser	Pro	Thr
	290					295					300				
Met	Ser	Arg	Pro	Arg	Ile	His	Phe	Leu	His	Pro	Ala	Tyr	Gln	Ser	Phe
305					310					315				320	
Thr	Asn	Asp	Ser	Val	Ser	Pro	Leu	Pro	Pro	Pro	Pro	Pro	Glu	Pro	Ala
			325						330					335	
Pro	Ala	Arg	Glu	Lys	Val	Glu	Asn	Gly	Gly	Pro	Thr	Thr	Phe	Lys	Met
		340						345					350		
Thr	Phe	Lys	Lys	Ala	Ala	Asn	Ile	Pro	Ile	Leu	Lys	Thr	Ser	Ala	Phe
		355					360					365			
Glu	Gln	Pro	Ser	Ser	Pro	Pro	Pro	Ser	Ser	Ser	Val	Ser	Ser	Ser	Ile
	370					375					380				
Ser	Leu	Ser	Glu	Val	Asn	Ser	Ser	Thr	Ser	Ile	Ala	Ser	Glu	Ser	Ser

385	Pro	Ala	Lys	Arg	Ser	390	Ser	Asn	Phe	Asp	Leu	395	Thr	Ala	Ser	Asn	Glu	Leu	400
					405						410							415	
Pro	Pro	Pro	Gln	Met	Val	Glu	Leu	Pro	Lys	Leu	Ser	Phe	Phe	Asn	Met				
			420						425						430				
Pro	Pro	Ala	Val	Arg	Ser	Ala	Glu	Asp	Asp	Ser	Ala	Met	Thr	Ser	Glu				
		435					440					445							
Glu	Pro	Ile	Leu	Leu	Leu	Arg	Ser	Pro	Asn	Ser	Ala	Thr	Pro	Asp	Asp				
	450					455					460								
Asp	Ala	Leu	Phe	Leu	Thr	Thr	Pro	Pro	Pro	Pro	Lys	Met	Thr	Glu	Ser				
465					470						475							480	
Glu	Ile	Gln	Ala	Leu	Lys	Val	Ala	Thr	Glu	Lys	Val	Asn	Gln	Val	Ile				
				485					490						495				
Ala	Arg	Arg	Glu	Asp	Ser	Glu	Lys	Asp	Val	Arg	His	Arg	Glu	Asp	Arg				
			500					505					510						
Asp	Asp	Tyr	Asp	Arg	Arg	Arg	Asp	Asp	Arg	Asp	Arg	Arg	Ser	Arg	Lys				
	515						520					525							
Thr	Asp	Ser	Glu	Arg	Asn	Asp	Gln	Arg	Gly	Arg	Gln	Arg	Glu	Asp	Asp				
	530					535					540								
Glu	Arg	Arg	Ala	Arg	Glu	Arg	Glu	Arg	Glu	Val	Thr	Lys	Arg	His	Asp				
545					550					555					560				
Arg	Glu	Arg	Glu	Glu	Met	Arg	Leu	Gln	Lys	Gln	Lys	Asp	Glu	Glu	Arg				
			565						570						575				
Arg	Lys	Lys	Asp	Glu	Glu	Glu	Arg	Ile	Gln	Lys	Glu	Asn	Asp	Glu	Lys				
			580					585					590						
Lys	Gln	Lys	Glu	Asp	Glu	Ala	Lys	Met	Glu	Glu	Glu	Lys	Lys	Lys	Ile				
	595						600					605							
Lys	Glu	Glu	Glu	Met	Lys	Ile	Pro	Glu	Phe	Glu	Leu	Ile	Ser	Glu	Ser				
	610					615					620								
Lys	Tyr	Leu	Thr	Arg	Asn	Ala	Asn	Lys	Lys	Lys	Thr	Glu	Ser	Leu	Thr				
625					630					635					640				
Cys	Glu	Cys	His	Arg	Thr	Gly	Gly	Asn	Cys	Ser	Asp	Asn	Thr	Cys	Val				
			645						650					655					
Asn	Arg	Ala	Met	Leu	Thr	Glu	Cys	Pro	Ser	Ser	Cys	Gln	Val	Lys	Cys				
		660						665					670						
Lys	Asn	Gln	Arg	Phe	Ala	Lys	Lys	Lys	Tyr	Ala	Ala	Val	Glu	Ala	Phe				
	675						680					685							
His	Thr	Gly	Thr	Ala	Lys	Gly	Cys	Gly	Leu	Arg	Ala	Val	Lys	Asp	Ile				
	690					695					700								
Lys	Lys	Gly	Arg	Phe	Ile	Ile	Glu	Tyr	Ile	Gly	Glu	Val	Val	Glu	Arg				
705					710					715					720				
Asp	Asp	Tyr	Glu	Lys	Arg	Lys	Thr	Lys	Tyr	Ala	Ala	Asp	Lys	Lys	His				
			725						730					735					
Lys	His	His	Tyr	Leu	Cys	Asp	Thr	Gly	Val	Tyr	Thr	Ile	Asp	Ala	Thr				
			740					745					750						
Val	Tyr	Gly	Asn	Pro	Ser	Arg	Phe	Val	Asn	His	Ser	Cys	Asp	Pro	Asn				
	755						760					765							
Ala	Ile	Cys	Glu	Lys	Trp	Ser	Val	Pro	Arg	Thr	Pro	Gly	Asp	Val	Asn				
	770					775						780							
Arg	Val	Gly	Phe	Phe	Ser	Lys	Arg	Phe	Ile	Lys	Ala	Gly	Glu	Glu	Ile				
785					790					795					800				
Thr	Phe	Asp	Tyr	Gln	Phe	Val	Asn	Tyr	Gly	Arg	Asp	Ala	Gln	Gln	Cys				
			805						810					815					
Phe	Cys	Gly	Ser	Ala	Ser	Cys	Ser	Gly	Trp	Ile	Gly	Gln	Lys	Pro	Glu				
			820					825					830						
Glu	Phe	Ser	Ser	Asp	Glu	Asp	Asp	Asp	Ile	Val	Thr	Thr	Arg	His	Ile				
	835						840					845							
Asn	Met	Asp	Glu	Glu	Glu	Glu	Glu	Lys	Leu	Glu	Gly	Leu	Asp	His	Leu				

850	855	860
Gly Asn His Glu Arg Asn Glu Val Ile Lys Asp Met Leu Asp Asp Leu		
865	870	875
Val Ile Arg Asn Lys Lys His Ala Arg Lys Val Ile Thr Ile Ala Ser		
	885	890
Ala Met Thr Asp Tyr Ser Gln Arg Val Asp Val Ile Gln Glu Ile Phe		
	900	905
Ser Ser Asp Thr Ser Val Thr Val Gln Lys Phe Tyr Ala Lys Glu Gly		
	915	920
Met Ala Thr Leu Met Ala Glu Trp Leu Ser Glu Asp Asp Tyr Ser Leu		
	930	935
Asp Asn Leu Lys Leu Val Gln Ala Ile Leu Lys Ala Leu His Thr Glu		
	945	950
Leu Phe Asp Ser Cys Ala Lys Asn Asp Arg Leu Leu Arg Asp Ser Thr		
	965	970
Ser Arg Trp Val Asn Ala Lys Met Asp Glu Tyr Val Asp Ile Gln Val		
	980	985
Ile Ala Asp Ser Leu Ile Ala Cys Val Glu Asp Pro Val Gln Glu Tyr		
	995	1000
Lys Asp Val Cys Lys Val Ile Glu Lys Gly Leu Val Glu Asn Phe Thr		
	1010	1015
Arg Ala Lys Glu Met Ala Tyr Arg Leu Asn Gln Tyr Trp Phe Asn Arg		
	1025	1030
Ser Val Ser Phe Lys Ile Pro Lys Lys Ile Arg Asp Pro Val Pro Lys		
	1045	1050
Asp Val Pro Val Arg Gln Glu Asp Ala Thr Thr Ser Ser Gln Ser His		
	1060	1065
Asp Asn Ser Ser Arg Thr Val Ser Pro Asn His Arg His His Ser Ser		
	1075	1080
Ser Tyr Ser Asn Ser Cys Tyr Gln Glu Arg Glu Pro Ser His Ile Arg		
	1090	1095
Phe Phe Asn Asn Gly Asn Asp Val His Gln Tyr Arg Phe Gly Gly Tyr		
	1105	1110
His Gly Asn Asn Tyr Asn Asp Asn Tyr Phe Ser Arg Arg Pro Asn Lys		
	1125	1130
Asp Ser Tyr Arg Asp Arg Arg Arg Phe Asn Gly Arg Arg Ser Arg Ser		
	1140	1145
Arg Ser Arg Ser Val Ser Pro Gln Asn Tyr Lys Arg Arg Lys Leu Asp		
	1155	1160
Glu His Asp Asn Asn His Arg Gln Arg Ser Pro Ile Arg Asp Arg His		
	1170	1175
Thr Ser Pro Gly Gly Glu Lys Thr Pro Ser Ser Asn Asn Ser Gly Glu		
	1185	1190
Arg Asn Tyr Lys Arg Leu Asp Ile Arg Gly Ala Arg Ile Lys Thr Ile		
	1205	1210
Lys Glu Asp Leu Glu Ala Ala Ala Ala Ala Ala Ala Ala Val		
	1220	1225
Pro Ser Glu Val Gln Ala Tyr Pro His Glu His Thr Ala Val His Gln		
	1235	1240
Ser Val Tyr Gln Met Pro Gly Tyr Glu Ser Tyr Gly Val Tyr Asp Pro		
	1250	1255
Val Asn Gly Val Tyr Met Tyr Pro His Pro Gly Ala Gly Tyr Tyr Pro		
	1265	1270
Pro Ala Tyr Pro Gln Gln Pro Ile Met Leu Thr Met Asp Thr Leu Pro		
	1285	1290
Pro Asn Asp Arg Leu Gly Glu Leu Tyr Glu Lys Ala Ser Ile Glu Gln		
	1300	1305
Leu Ala Gln Arg Asp Ala Ile Val Arg Gln Glu Leu Glu Leu Ile Arg		

1315	1320	1325
Ile Gln Ile Glu Arg Lys	Thr Ala Gln Lys Glu	Ala Ile Lys Ala Ala
1330	1335	1340
Cys Arg Arg Ala Asn Glu	Glu Glu Ala Lys Arg	Gln Glu Ala Leu Ala
1345	1350	1355
Lys Thr Lys Tyr Val Trp	Ala Ile Ala Lys Ser	Glu Ala Gly Glu Thr
1365	1370	1375
Tyr Tyr Tyr Asn Lys Ile	Thr Lys Glu Thr Gln	Trp Thr Ala Pro Thr
1380	1385	1390
Pro Val Gln Gly Leu Leu	Glu Pro Ala Cys Gly	Ala Ser Pro Asp Thr
1395	1400	1405
Thr Val Val Ile Ala Asp	Glu Ile Thr Glu Glu	Glu Gln Gln Ala Glu
1410	1415	1420
Val Leu Glu Lys Pro Arg	Val Val Lys Glu Glu	Val Ile Glu Pro Gly
1425	1430	1435
Ser Gln Ser Glu Thr Gln	Lys Glu Ser Pro Glu	Lys Val Arg Val Val
1445	1450	1455
Val Pro Lys Val Glu Val	Glu Arg Ser Pro Ser	Pro Lys Ser Ser Arg
1460	1465	1470
Asp Arg Glu Lys Asp Arg	Glu Lys Ser Arg Glu	Lys Asp Arg Glu Arg
1475	1480	1485
Asp Arg Asp Arg Arg Glu	Gly Ser Lys His Arg	Asp Ser Tyr His Gly
1490	1495	1500
His Arg Asn Gly Ser Ser	Ser Val Ser Glu Arg	Arg Met Arg Glu Phe
1505	1510	1515
Lys His Glu Leu Glu Arg	Ser Thr Arg Ser Ala	Val Arg Ser Arg Leu
1525	1530	1535
Gln His Gln Arg Asp Ala	Ser Ser Asp Lys Thr	Thr Trp Leu Ile Lys
1540	1545	1550
Leu Ile Tyr Arg Glu Ile	Phe Lys Arg Glu Ser	Ala Gln Ser Gly Phe
1555	1560	1565
Asp Tyr Arg Phe Ser Glu	Asn Thr Asp Lys Lys	Val Lys Asn Tyr Val
1570	1575	1580
Lys Ser Tyr Ile Asp Arg	Lys Leu Glu Ser Asn	Asp Leu Trp Lys Glu
1585	1590	1595
Tyr Ser Arg Pro		1600

<210> 26
 <211> 7333
 <212> DNA
 <213> Caenorhabditis elegans

<220>
 <221> CDS
 <222> (1001)...(1096)

<221> CDS
 <222> (1166)...(1453)

<221> CDS
 <222> (1501)...(2199)

<221> CDS
 <222> (2298)...(2730)

<221> CDS

<222> (3234) ... (3847)

<221> CDS

<222> (4148) ... (5778)

<221> CDS

<222> (6111) ... (6333)

<400> 26

```

gcttgcatcg aaactcttct cattattttac gtgatgatca catctttcgt tgggctgtac 60
tcccttccgg ttcttcgttc tcttcgacct gttcgaaaag atactccaat gccaacgata 120
attattaatt cttcaatagt tcttggtgtt gcatccgctc tcccagtagc tgttaacaca 180
gttggaatga caacttttga tcttctcggc tcccactcat cgctccaatg gcttgatga 240
tttcgagtcg ttgttgccct taatactcta ttcgtcgtgt tgtctgtcgc atttctcttc 300
aatcaattga ctgcttcaat gagaaggcaa atctggaagt ggtaagctgt gcaatttaaa 360
gtttaaattc ttattaattt ttttgcagga tatgtcaact acgatgtgga atcagacggg 420
agagtgatgc ggatgaaacc attgagatcc ttagaggcga taagaaaagc aattgaattt 480
ctttcctttt tcaacacttc ttacccatgt tcatcatttt aatcttttca ttacaaaaac 540
aaggtcctat tttttttctc ggggtactact cgctttttct aataattcag aatcatcaat 600
ttttgccaac ctctagcttt acatgtctgt ttttcatcat tttctctcaa gcattctcct 660
aatatattat gttccctagt atttccctc agtcagcaat tttctcgtcg tcgaaaccgt 720
ttagctttac tttcaatcaa aacgtggaac atttttcaaa ctatttgaag ccaaaaaaaaa 780
ccagggtctt tgtatatgta ccatattttc cctctgattt tctttatcgc cttctctttt 840
catgtagaat aactgaaata caaaccattt taattttttc ttttaattat caatactgtc 900
cgtataggta aaaattattt cttcagggtt gaaaaaatcc gaaatatgta tctgcaactc 960
ttcagggcat tgcctcaatt aatttttatc taatattcag atg gac caa caa gaa 1015
                                     Met Asp Gln Gln Glu
                                     1               5

```

```

cca tcg aat aac gta gat acg agc agt att ctt tcg gat gat ggg atg 1063
Pro Ser Asn Asn Val Asp Thr Ser Ser Ile Leu Ser Asp Asp Gly Met
                10                15                20

```

```

gaa aca cag gaa caa agt tca ttc gtc act gct gtgagtgaat ttatttataaa 1116
Glu Thr Gln Glu Gln Ser Ser Phe Val Thr Ala
                25                30

```

```

tttcgcttcg gagattcatt gtcataataat tcaatttatc gattttcag aca att gac 1174
                                     Thr Ile Asp
                                     35

```

```

cta aca gtg gac gac tac gat gaa aca gaa ata cag gag att ctg gat 1222
Leu Thr Val Asp Asp Tyr Asp Glu Thr Glu Ile Gln Glu Ile Leu Asp
                40                45                50

```

```

aat gga aaa gca gaa gaa gga aca gat gaa gat tct gat tta gtt gaa 1270
Asn Gly Lys Ala Glu Glu Gly Thr Asp Glu Asp Ser Asp Leu Val Glu
                55                60                65

```

```

ggg att ctt aac gct aat tca gat gtc caa gcg ctc ctt gat gcg cca 1318
Gly Ile Leu Asn Ala Asn Ser Asp Val Gln Ala Leu Leu Asp Ala Pro
                70                75                80

```

```

tct gag caa gta gct caa gct ctt aat tcg ttc ttc gga aat gag agt 1366
Ser Glu Gln Val Ala Gln Ala Leu Asn Ser Phe Phe Gly Asn Glu Ser
                85                90                95

```

```

gaa caa gaa gct gtt gca gca caa aga cgg gtt gat gcg gag aag act 1414

```

Glu Gln Glu Ala Val Ala Ala Gln Arg Arg Val Asp Ala Glu Lys Thr	
100 105 110 115	
gcc aaa gat gaa gct gaa ctc aag caa cag gaa gag gcg gtttagattgc	1463
Ala Lys Asp Glu Ala Glu Leu Lys Gln Gln Glu Glu Ala	
120 125	
aataaaggaa acaataataa aattatttta ttttcag gaa gat ctt att ata gaa	1518
Glu Asp Leu Ile Ile Glu	
130	
gat tcg ata gtc aaa act gat gaa gaa aaa caa gca gtt cga aga ctg	1566
Asp Ser Ile Val Lys Thr Asp Glu Glu Lys Gln Ala Val Arg Arg Leu	
135 140 145 150	
aaa atc aac gaa ttt tta tcg tgg ttc aca agg ctc ctt cca gaa caa	1614
Lys Ile Asn Glu Phe Leu Ser Trp Phe Thr Arg Leu Leu Pro Glu Gln	
155 160 165	
ttt aaa aat ttc gaa ttc aca aat ccg aac tat ctg aca gaa tct atc	1662
Phe Lys Asn Phe Glu Phe Thr Asn Pro Asn Tyr Leu Thr Glu Ser Ile	
170 175 180	
agc gat tca ccg gtt gta aat gtc gat aaa tgc aag gaa att gtc aaa	1710
Ser Asp Ser Pro Val Val Asn Val Asp Lys Cys Lys Glu Ile Val Lys	
185 190 195	
tcg ttc aag gaa agt gaa tca ctt gag gga ctt tca cag aaa tac gaa	1758
Ser Phe Lys Glu Ser Glu Ser Leu Glu Gly Leu Ser Gln Lys Tyr Glu	
200 205 210	
tta att gat gaa gac gtg cta gtc gct gct att tgt att ggc gtt ctc	1806
Leu Ile Asp Glu Asp Val Leu Val Ala Ala Ile Cys Ile Gly Val Leu	
215 220 225 230	
gat acc aac aac gaa gaa gat gtc gac ttt aat gtt cta tgt gat gat	1854
Asp Thr Asn Asn Glu Glu Asp Val Asp Phe Asn Val Leu Cys Asp Asp	
235 240 245	
cgt atc gac gat tgg agt ata gaa aaa tgt gtc act ttt ctt gat tat	1902
Arg Ile Asp Asp Trp Ser Ile Glu Lys Cys Val Thr Phe Leu Asp Tyr	
250 255 260	
cca aat act gga ttg aat tcg aaa aat gga ccg ttg aga ttc atg cag	1950
Pro Asn Thr Gly Leu Asn Ser Lys Asn Gly Pro Leu Arg Phe Met Gln	
265 270 275	
ttt act gtc aca tca cct gca tca gca att ctc atg ctc act ctg att	1998
Phe Thr Val Thr Ser Pro Ala Ser Ala Ile Leu Met Leu Thr Leu Ile	
280 285 290	
cga tta cgc gaa gaa ggg cat ccg tgt cga tta gat ttt gat tca aat	2046
Arg Leu Arg Glu Glu Gly His Pro Cys Arg Leu Asp Phe Asp Ser Asn	
295 300 305 310	
ccg act gat gat tta ctc ttg aat ttc gat caa gtg gaa ttt tct aat	2094
Pro Thr Asp Asp Leu Leu Leu Asn Phe Asp Gln Val Glu Phe Ser Asn	
315 320 325	

aat atc att gat acg gca gtc aaa tac tgg gat gat cag aag gaa aac 2142
 Asn Ile Ile Asp Thr Ala Val Lys Tyr Trp Asp Asp Gln Lys Glu Asn
 330 335 340

ggt gcg cag gat aaa att ggc agg cga gta tta atc aaa ctc aca act 2190
 Gly Ala Gln Asp Lys Ile Gly Arg Arg Val Leu Ile Lys Leu Thr Thr
 345 350 355

gtt ttg aaa gtattttcat aattatcact taaatacctt ttagagagct 2239
 Val Leu Lys
 360

caacgacttc ttccacgaaa tcgagtcaac atcagcagaa ttcaaacaac attttgag 2297
 aac gcc gtt ggc agc cgt aat gaa ata att caa ctt gtc aac gag aaa 2345
 Asn Ala Val Gly Ser Arg Asn Glu Ile Ile Gln Leu Val Asn Glu Lys
 365 370 375

att ccc gat ttt gat ggc act gag gct gct gtg aat gag agt ttt aca 2393
 Ile Pro Asp Phe Asp Gly Thr Glu Ala Ala Val Asn Glu Ser Phe Thr
 380 385 390

tcc gat caa cga acc gaa att atc aac tct cgt gca ata atg gag aca 2441
 Ser Asp Gln Arg Thr Glu Ile Ile Asn Ser Arg Ala Ile Met Glu Thr
 395 400 405

tta aaa gcc gag atg aag ctc gcc atc gcc gaa gct cag aaa gtt tac 2489
 Leu Lys Ala Glu Met Lys Leu Ala Ile Ala Glu Ala Gln Lys Val Tyr
 410 415 420 425

gac acc aag act gac ttc gaa aaa ttc ttc gtt ttg aca gtt gga gat 2537
 Asp Thr Lys Thr Asp Phe Glu Lys Phe Phe Val Leu Thr Val Gly Asp
 430 435 440

ttc tgt ctg gct cgc gcc aat cct tct gac gat gca gaa tta aca tac 2585
 Phe Cys Leu Ala Arg Ala Asn Pro Ser Asp Asp Ala Glu Leu Thr Tyr
 445 450 455

gcc ata gtt cag gat cgt gtg gat gca atg acc tat aag gtt aaa ttt 2633
 Ala Ile Val Gln Asp Arg Val Asp Ala Met Thr Tyr Lys Val Lys Phe
 460 465 470

atc gac aca agt cag atc aga gag tgt aac atc aga gat tta gcc atg 2681
 Ile Asp Thr Ser Gln Ile Arg Glu Cys Asn Ile Arg Asp Leu Ala Met
 475 480 485

act acg cag gga atg tat gac ccg agt ttg aat aca ttt ggt gat gtt 2729
 Thr Thr Gln Gly Met Tyr Asp Pro Ser Leu Asn Thr Phe Gly Asp Val
 490 495 500 505

g gtgagtttta agttaaaatt gatatttaatt attacatctg ttatgtagaa 2780
 taagggtttc gggtttttcga ttttattaga aaatcgaaaa ttttagtttt tgtgttaaat 2840
 ttaaaaaaat caaaatttga ttcactatca agtccgtttt tctcttctca aaattgacaa 2900
 aattttgata atctagaatt ttcgtcccgat atatttttca acgaaaaacc atttaaaatt 2960
 ttccatgatt ggattttcgg ttgatctaga aaaaaatggg gctaaacact aaatttgaaa 3020
 aagtttgaaa caaattcaaa tccaaatatt tcatgaaaaa cttgtaaaaat atattatgta 3080
 cacaaaaaaa cgtttcaagt gtagcagttg ttttttggtg tcccaaaaaa gcagatgttt 3140
 gtcagaatcc attaaacaac aaaaaaatcc aaaaactcaa cctggcctag atatcagttt 3200

catgatcgaa gtatctaaaa tcattgtttt cag gt ctt cga gtt gcc tgt cgc	3253
Gly Leu Arg Val Ala Cys Arg	
510	
caa gtt att tcc tcc agc caa ttt gga aaa aaa aca att tgg ctt acc	3301
Gln Val Ile Ser Ser Ser Gln Phe Gly Lys Lys Thr Ile Trp Leu Thr	
515 520 525	
ggt aca gct gcc gga cgt cgc aga gct cat aga tcc gat ttt cta att	3349
Gly Thr Ala Ala Gly Arg Arg Arg Ala His Arg Ser Asp Phe Leu Ile	
530 535 540	
ttc ttc gac aac gga acc gat gca tac gtg tca gct ccg aca atg cct	3397
Phe Phe Asp Asn Gly Thr Asp Ala Tyr Val Ser Ala Pro Thr Met Pro	
545 550 555 560	
ggt gaa cca ggt tat gaa gtt gct tct gaa aag aaa agt gta ttt tct	3445
Gly Glu Pro Gly Tyr Glu Val Ala Ser Glu Lys Lys Ser Val Phe Ser	
565 570 575	
ctc aaa gaa atg att gcg aag atg aat gct gct cag att gct att atg	3493
Leu Lys Glu Met Ile Ala Lys Met Asn Ala Ala Gln Ile Ala Ile Met	
580 585 590	
gtt gga cag cca gta gga aag gaa gga aat ctg gat tat ttt ttg aca	3541
Val Gly Gln Pro Val Gly Lys Glu Gly Asn Leu Asp Tyr Phe Leu Thr	
595 600 605	
ttt cat tgg att cga caa tct cac aga tca gcg tat att ccg gat ttt	3589
Phe His Trp Ile Arg Gln Ser His Arg Ser Ala Tyr Ile Arg Asp Phe	
610 615 620	
atg aaa gaa ttt ccg gaa tgg cca ctt ctc aag atg cca gtt gga atg	3637
Met Lys Glu Phe Pro Glu Trp Pro Leu Leu Lys Met Pro Val Gly Met	
625 630 635 640	
cga atc tgt ttg tac aat tct ctt gtt gat cga cgt aag aaa atg gtg	3685
Arg Ile Cys Leu Tyr Asn Ser Leu Val Asp Arg Arg Lys Lys Met Val	
645 650 655	
aca gtg att gga act gat cga gct ttt gct att gtg aga cac gaa gca	3733
Thr Val Ile Gly Thr Asp Arg Ala Phe Ala Ile Val Arg His Glu Ala	
660 665 670	
ccg aat cca ttg gct cct ggg aat aga tgt aca gac ttt ccg tgc aat	3781
Pro Asn Pro Leu Ala Pro Gly Asn Arg Cys Thr Asp Phe Pro Cys Asn	
675 680 685	
gat aga aat cat cag cat att gac gag aaa atc tat aga gga tct cat	3829
Asp Arg Asn His Gln His Ile Asp Glu Lys Ile Tyr Arg Gly Ser His	
690 695 700	
aga ttg gaa ggc gca gcg gtaagatttt atttgaaaaa ttgatacaaa	3877
Arg Leu Glu Gly Ala Ala	
705 710	
acgaggattt tctaaaatta ttttattttt atttgatttg atttcttata attgataatc	3937
aagggtttttt ggatgttttg ttagagaaat cgaaaagggg aacttccaaa aaaaagctgt	3997

gaaatcaatt tttgctttta ataatatcca agtttcatct tcaaagtttt ttctataaaa 4057
 tggacacaaa cttttcaacg ttttcaaaaa aaaggttccg aaaatatgaa aaaaggagaa 4117
 agaaatcatg aaaattttgt attatttcag cac aag aag cac atg atc tcg aca 4171
 His Lys Lys His Met Ile Ser Thr
 715

aat aac aat ctg tcg caa cgc aga aaa gac cag ctt caa tca cag ttc 4219
 Asn Asn Asn Leu Ser Gln Arg Arg Lys Asp Gln Leu Gln Ser Gln Phe
 720 725 730

gag cca acc gac atg att cgt tcg atg cca gag agg aat cac caa caa 4267
 Glu Pro Thr Asp Met Ile Arg Ser Met Pro Glu Arg Asn His Gln Gln
 735 740 745 750

gtc gtt aaa aag aaa acg acg ggc acc aat cag aat gtc gct tcg aca 4315
 Val Val Lys Lys Lys Thr Thr Gly Thr Asn Gln Asn Val Ala Ser Thr
 755 760 765

aat gat gca aaa tcg aag aga gaa att gaa ata aga aag aaa aat caa 4363
 Asn Asp Ala Lys Ser Lys Arg Glu Ile Glu Ile Arg Lys Lys Asn Gln
 770 775 780

ttc tta ttt aac aag att att gtt cca ata ccc gtc cta aca cca ttg 4411
 Phe Leu Phe Asn Lys Ile Ile Val Pro Ile Pro Val Leu Thr Pro Leu
 785 790 795

gaa aat ctc aag gct cat gct caa tgt ggt cca gat tgt cta cag aaa 4459
 Glu Asn Leu Lys Ala His Ala Gln Cys Gly Pro Asp Cys Leu Gln Lys
 800 805 810

atg gat gcg gat ccg tat gaa gca aga ttc cat cga aat tca cca ata 4507
 Met Asp Ala Asp Pro Tyr Glu Ala Arg Phe His Arg Asn Ser Pro Ile
 815 820 825 830

cat act cct ctt ttg tgt ggt tgg aga cga att atg tac aca atg agt 4555
 His Thr Pro Leu Leu Cys Gly Trp Arg Arg Ile Met Tyr Thr Met Ser
 835 840 845

act gga aag aag cgg gga gca gtg aag aaa aac att att tac ttt tct 4603
 Thr Gly Lys Lys Arg Gly Ala Val Lys Lys Asn Ile Ile Tyr Phe Ser
 850 855 860

cca tgc gga gcc gct ctt cac cag atc agc gac gtc tct gaa tat att 4651
 Pro Cys Gly Ala Ala Leu His Gln Ile Ser Asp Val Ser Glu Tyr Ile
 865 870 875

cat gtc acc aga agt tta ttg acg att gat tgt ttt tca ttt gat gca 4699
 His Val Thr Arg Ser Leu Leu Thr Ile Asp Cys Phe Ser Phe Asp Ala
 880 885 890

cga atc gat act gcc act tat att act gtt gac gat aaa tat ttg aag 4747
 Arg Ile Asp Thr Ala Thr Tyr Ile Thr Val Asp Asp Lys Tyr Leu Lys
 895 900 905 910

gtt gct gat ttt tcg ctt gga acc gaa gga atc cca att cca cta gtg 4795
 Val Ala Asp Phe Ser Leu Gly Thr Glu Gly Ile Pro Ile Pro Leu Val
 915 920 925

aac agc gtg gat aac gat gag cct cca tca ttg gaa tat tcg aaa cga Asn Ser Val Asp Asn Asp Glu Pro Pro Ser Leu Glu Tyr Ser Lys Arg 930 935 940	4843
cga ttc caa tac aat gat caa gtg gat ata tcg agt gtt agc cga gat Arg Phe Gln Tyr Asn Asp Gln Val Asp Ile Ser Ser Val Ser Arg Asp 945 950 955	4891
ttc tgt tct gga tgc tct tgt gat ggt gat tgc agt gac gca tcg aag Phe Cys Ser Gly Cys Ser Cys Asp Gly Asp Cys Ser Asp Ala Ser Lys 960 965 970	4939
tgt gaa tgc caa caa ttg tcc att gaa gca atg aaa cga ctc ccc cat Cys Glu Cys Gln Gln Leu Ser Ile Glu Ala Met Lys Arg Leu Pro His 975 980 985 990	4987
aat tta caa ttc gac gga cac gac gaa ttg tat gag agt tca gaa aaa Asn Leu Gln Phe Asp Gly His Asp Glu Leu Tyr Glu Ser Ser Glu Lys 995 1000 1005	5035
caa aat aaa ttt tta aaa cta ttt ttt ttc aga gtt cct cac tat caa Gln Asn Lys Phe Leu Lys Leu Phe Phe Phe Arg Val Pro His Tyr Gln 1010 1015 1020	5083
aat cgt ctt ctc agc agt aag gtt atc agt gga ctc tat gaa tgc aac Asn Arg Leu Leu Ser Ser Lys Val Ile Ser Gly Leu Tyr Glu Cys Asn 1025 1030 1035	5131
gat cag tgt tca tgc cat cga aag tct tgt tac aac aga gtt gtt cag Asp Gln Cys Ser Cys His Arg Lys Ser Cys Tyr Asn Arg Val Val Gln 1040 1045 1050	5179
aac aat atc aag tat cct atg cat gtg agt tta ttt aac gat gat aca Asn Asn Ile Lys Tyr Pro Met His Val Ser Leu Phe Asn Asp Asp Thr 1055 1060 1065 1070	5227
tac caa tta ttg ttt ttt ctt cag atc ttc aaa act gct caa tcc gga Tyr Gln Leu Leu Phe Phe Leu Gln Ile Phe Lys Thr Ala Gln Ser Gly 1075 1080 1085	5275
tgg gga gtc cga gct ttg acg gat att cct caa agt acg ttc att tgc Trp Gly Val Arg Ala Leu Thr Asp Ile Pro Gln Ser Thr Phe Ile Cys 1090 1095 1100	5323
acg tat gta ggt gct ata ctg acg gat gat ttg gct gat gaa cta aga Thr Tyr Val Gly Ala Ile Leu Thr Asp Asp Leu Ala Asp Glu Leu Arg 1105 1110 1115	5371
aat gcg gat caa tac ttc gct gat ttg gac ttg aag gat acc gtg gag Asn Ala Asp Gln Tyr Phe Ala Asp Leu Asp Leu Lys Asp Thr Val Glu 1120 1125 1130	5419
ctg gaa aag ggt cgc gaa gat cat gaa act gat ttt ggt tac gga gga Leu Glu Lys Gly Arg Glu Asp His Glu Thr Asp Phe Gly Tyr Gly Gly 1135 1140 1145 1150	5467
gac gag tca gat tat gat gac gaa gaa gga agt gat ggt gac tcc ggt Asp Glu Ser Asp Tyr Asp Asp Glu Glu Gly Ser Asp Gly Asp Ser Gly	5515

1155	1160	1165	
gat gat gta atg aac aaa atg gtg aaa cgt caa gac tct tcg gag agt Asp Asp Val Met Asn Lys Met Val Lys Arg Gln Asp Ser Ser Glu Ser 1170 1175 1180			5563
ggt gaa gaa aca aaa cgg ctg aca aga cag aaa aga aag caa tct aaa Gly Glu Glu Thr Lys Arg Leu Thr Arg Gln Lys Arg Lys Gln Ser Lys 1185 1190 1195			5611
aaa tcc ggt aaa gga gga agt gtg gag aaa gat gac acc act cca aga Lys Ser Gly Lys Gly Gly Ser Val Glu Lys Asp Asp Thr Thr Pro Arg 1200 1205 1210			5659
gat tca atg gaa aag gat aat att gaa agt aaa gac gaa ccc gtt ttc Asp Ser Met Glu Lys Asp Asn Ile Glu Ser Lys Asp Glu Pro Val Phe 1215 1220 1225 1230			5707
aat tgg gat aag tat ttt gag cgg ttt cca ttg tat gtt ata gat gca Asn Trp Asp Lys Tyr Phe Glu Pro Phe Pro Leu Tyr Val Ile Asp Ala 1235 1240 1245			5755
aaa cag aga gga aat ctt gga ag gtaagatcac aattttattc attaaaaaaa Lys Gln Arg Gly Asn Leu Gly Arg 1250			5808
tttttttagag attttgcttt aaatgataaa aaatggacaa accaaccggt tgctctttct tttggtttat caacctttct ctatggaaaa aattctgaaa aattaacaaa cagtatttca cggtgaaaag tgaagaaaaa agcaaaaaaa ggaaacaaat ttcaaaacgg ttctactcca tcttaaaaaa actaaaattc gtaaaaagtc atttggatg ttttggagac tataatacaa ttgagaaaaat ttgaaaaacc ggcactccaa agatacaatc ataaattttc gataactttc ag a ttc ttg aat cac tct tgc gat cgg aat gtg cac gtt caa cac gtc Phe Leu Asn His Ser Cys Asp Pro Asn Val His Val Gln His Val 1255 1260 1265			5868 5928 5988 6048 6108 6156
atg tac gat acg cat gat ctt cgt ctt cca tgg gtc gcg ttt ttc aca Met Tyr Asp Thr His Asp Leu Arg Leu Pro Trp Val Ala Phe Phe Thr 1270 1275 1280 1285			6204
cga aaa tac gtg aaa gcc ggc gat gag cta acc tgg gac tat caa tat Arg Lys Tyr Val Lys Ala Gly Asp Glu Leu Thr Trp Asp Tyr Gln Tyr 1290 1295 1300			6252
act caa gat cag acg gct acc aca caa ctc aca tgc cac tgc gga gct Thr Gln Asp Gln Thr Ala Thr Thr Gln Leu Thr Cys His Cys Gly Ala 1305 1310 1315			6300
gaa aac tgc acc ggc cgt ttg ctg aaa agt taa agaattgttg ttatttcctt Glu Asn Cys Thr Gly Arg Leu Leu Lys Ser *			6353
cccagttatg ttttcctttt tttttaagta tttattttatt tatttaattt ttattttggt tattgttcaa tcgttttaaaa tctccctttg aaaacagcat ctcatatgta tgatctaaac acgtattttac ctcgtaaggg tttgccaaat agtttctttg gttttcattt tgattttctc tgcaataaaa atgttttaaa aaagacatta ttttttttaa tagtcagtac agttttgatg tctccaatct atttcagttt acaatttttaa aatatagaat atatatatatt aggtttcata agttatgcat cgattacggg ttctacgtca cttgaagttc tgcattttcca cgtcacatag gactactgta gtttttaaaaa atactcgttc attttgtaat aatatttcctt ctactagttt			6413 6473 6533 6593 6653 6713 6773

```

tgcttctggt aataatcgaa tttcaaaact ttagctaaaa tatttctttt tgaagaggct 6833
gcagcaaaat atgaaaagaa aagtccaact gaacatgtat tacttcgacc cgatacatat 6893
attggagggtg tcgccatgcg agaagatcaa attatttggc tcagagactc agaaaataga 6953
aaaatgattg caaaagaagt cacttatcca cctggattat tgaagatttt cgatgagatt 7013
ctagtgaatg cggctgataa taaagcaaga gattccagta tgaatcgggt ggaagtatgg 7073
ttagataggt aaatatattg caggaattta tgttctgcga caaagctacg atacgctgtc 7133
tcgccacgac aattgttttg gtaaatgcat gaaaatcgac gtgcaccttt aaataatact 7193
gtagtttttaa attctcgttt cttcaatttt tcataaatgg ttttccgatg aatatatgat 7253
tttaaaaaaa tctaaaattc acattaattt ataagaaaca aaattcctca aaaacgaaag 7313
tttggcgata cagtactatc                                     7333

```

<210> 27

<211> 1327

<212> PRT

<213> *Caenorhabditis elegans*

<400> 27

```

Met Asp Gln Gln Glu Pro Ser Asn Asn Val Asp Thr Ser Ser Ile Leu
 1          5          10          15
Ser Asp Asp Gly Met Glu Thr Gln Glu Gln Ser Ser Phe Val Thr Ala
 20          25          30
Thr Ile Asp Leu Thr Val Asp Asp Tyr Asp Glu Thr Glu Ile Gln Glu
 35          40          45
Ile Leu Asp Asn Gly Lys Ala Glu Glu Gly Thr Asp Glu Asp Ser Asp
 50          55          60
Leu Val Glu Gly Ile Leu Asn Ala Asn Ser Asp Val Gln Ala Leu Leu
 65          70          75          80
Asp Ala Pro Ser Glu Gln Val Ala Gln Ala Leu Asn Ser Phe Phe Gly
 85          90          95
Asn Glu Ser Glu Gln Glu Ala Val Ala Ala Gln Arg Arg Val Asp Ala
100          105          110
Glu Lys Thr Ala Lys Asp Glu Ala Glu Leu Lys Gln Gln Glu Glu Ala
115          120          125
Glu Asp Leu Ile Ile Glu Asp Ser Ile Val Lys Thr Asp Glu Glu Lys
130          135          140
Gln Ala Val Arg Arg Leu Lys Ile Asn Glu Phe Leu Ser Trp Phe Thr
145          150          155          160
Arg Leu Leu Pro Glu Gln Phe Lys Asn Phe Glu Phe Thr Asn Pro Asn
165          170          175
Tyr Leu Thr Glu Ser Ile Ser Asp Ser Pro Val Val Asn Val Asp Lys
180          185          190
Cys Lys Glu Ile Val Lys Ser Phe Lys Glu Ser Glu Ser Leu Glu Gly
195          200          205
Leu Ser Gln Lys Tyr Glu Leu Ile Asp Glu Asp Val Leu Val Ala Ala
210          215          220
Ile Cys Ile Gly Val Leu Asp Thr Asn Asn Glu Glu Asp Val Asp Phe
225          230          235          240
Asn Val Leu Cys Asp Asp Arg Ile Asp Asp Trp Ser Ile Glu Lys Cys
245          250          255
Val Thr Phe Leu Asp Tyr Pro Asn Thr Gly Leu Asn Ser Lys Asn Gly
260          265          270
Pro Leu Arg Phe Met Gln Phe Thr Val Thr Ser Pro Ala Ser Ala Ile
275          280          285
Leu Met Leu Thr Leu Ile Arg Leu Arg Glu Glu Gly His Pro Cys Arg
290          295          300
Leu Asp Phe Asp Ser Asn Pro Thr Asp Asp Leu Leu Leu Asn Phe Asp
305          310          315          320
Gln Val Glu Phe Ser Asn Asn Ile Ile Asp Thr Ala Val Lys Tyr Trp

```

				325					330					335			
Asp	Asp	Gln	Lys	Glu	Asn	Gly	Ala	Gln	Asp	Lys	Ile	Gly	Arg	Arg	Val		
			340						345					350			
Leu	Ile	Lys	Leu	Thr	Thr	Val	Leu	Lys	Asn	Ala	Val	Gly	Ser	Arg	Asn		
		355							360					365			
Glu	Ile	Ile	Gln	Leu	Val	Asn	Glu	Lys	Ile	Pro	Asp	Phe	Asp	Gly	Thr		
		370							375					380			
Glu	Ala	Ala	Val	Asn	Glu	Ser	Phe	Thr	Ser	Asp	Gln	Arg	Thr	Glu	Ile		
385						390					395				400		
Ile	Asn	Ser	Arg	Ala	Ile	Met	Glu	Thr	Leu	Lys	Ala	Glu	Met	Lys	Leu		
				405						410					415		
Ala	Ile	Ala	Glu	Ala	Gln	Lys	Val	Tyr	Asp	Thr	Lys	Thr	Asp	Phe	Glu		
				420						425					430		
Lys	Phe	Phe	Val	Leu	Thr	Val	Gly	Asp	Phe	Cys	Leu	Ala	Arg	Ala	Asn		
		435							440					445			
Pro	Ser	Asp	Asp	Ala	Glu	Leu	Thr	Tyr	Ala	Ile	Val	Gln	Asp	Arg	Val		
		450					455							460			
Asp	Ala	Met	Thr	Tyr	Lys	Val	Lys	Phe	Ile	Asp	Thr	Ser	Gln	Ile	Arg		
465						470					475				480		
Glu	Cys	Asn	Ile	Arg	Asp	Leu	Ala	Met	Thr	Thr	Gln	Gly	Met	Tyr	Asp		
				485						490					495		
Pro	Ser	Leu	Asn	Thr	Phe	Gly	Asp	Val	Gly	Leu	Arg	Val	Ala	Cys	Arg		
			500						505					510			
Gln	Val	Ile	Ser	Ser	Ser	Gln	Phe	Gly	Lys	Lys	Thr	Ile	Trp	Leu	Thr		
		515							520					525			
Gly	Thr	Ala	Ala	Gly	Arg	Arg	Arg	Ala	His	Arg	Ser	Asp	Phe	Leu	Ile		
		530					535					540					
Phe	Phe	Asp	Asn	Gly	Thr	Asp	Ala	Tyr	Val	Ser	Ala	Pro	Thr	Met	Pro		
545						550					555				560		
Gly	Glu	Pro	Gly	Tyr	Glu	Val	Ala	Ser	Glu	Lys	Lys	Ser	Val	Phe	Ser		
				565						570					575		
Leu	Lys	Glu	Met	Ile	Ala	Lys	Met	Asn	Ala	Ala	Gln	Ile	Ala	Ile	Met		
			580						585					590			
Val	Gly	Gln	Pro	Val	Gly	Lys	Glu	Gly	Asn	Leu	Asp	Tyr	Phe	Leu	Thr		
		595						600					605				
Phe	His	Trp	Ile	Arg	Gln	Ser	His	Arg	Ser	Ala	Tyr	Ile	Arg	Asp	Phe		
		610					615					620					
Met	Lys	Glu	Phe	Pro	Glu	Trp	Pro	Leu	Leu	Lys	Met	Pro	Val	Gly	Met		
625						630					635				640		
Arg	Ile	Cys	Leu	Tyr	Asn	Ser	Leu	Val	Asp	Arg	Arg	Lys	Lys	Met	Val		
				645						650					655		
Thr	Val	Ile	Gly	Thr	Asp	Arg	Ala	Phe	Ala	Ile	Val	Arg	His	Glu	Ala		
			660						665					670			
Pro	Asn	Pro	Leu	Ala	Pro	Gly	Asn	Arg	Cys	Thr	Asp	Phe	Pro	Cys	Asn		
		675						680					685				
Asp	Arg	Asn	His	Gln	His	Ile	Asp	Glu	Lys	Ile	Tyr	Arg	Gly	Ser	His		
		690					695				700						
Arg	Leu	Glu	Gly	Ala	Ala	His	Lys	Lys	His	Met	Ile	Ser	Thr	Asn	Asn		
705						710					715				720		
Asn	Leu	Ser	Gln	Arg	Arg	Lys	Asp	Gln	Leu	Gln	Ser	Gln	Phe	Glu	Pro		
				725						730					735		
Thr	Asp	Met	Ile	Arg	Ser	Met	Pro	Glu	Arg	Asn	His	Gln	Gln	Val	Val		
			740						745					750			
Lys	Lys	Lys	Thr	Thr	Gly	Thr	Asn	Gln	Asn	Val	Ala	Ser	Thr	Asn	Asp		
		755						760					765				
Ala	Lys	Ser	Lys	Arg	Glu	Ile	Glu	Ile	Arg	Lys	Lys	Asn	Gln	Phe	Leu		
		770					775					780					
Phe	Asn	Lys	Ile	Ile	Val	Pro	Ile	Pro	Val	Leu	Thr	Pro	Leu	Glu	Asn		

785		790		795		800
Leu Lys Ala His	Ala Gln Cys Gly Pro Asp	Cys Leu Gln Lys Met Asp				
	805	810			815	
Ala Asp Pro Tyr	Glu Ala Arg Phe His Arg	Asn Ser Pro Ile His Thr				
	820	825			830	
Pro Leu Leu Cys	Gly Trp Arg Arg Ile Met Tyr	Thr Met Ser Thr Gly				
	835	840			845	
Lys Lys Arg Gly	Ala Val Lys Lys Asn Ile Ile	Tyr Phe Ser Pro Cys				
	850	855			860	
Gly Ala Ala Leu	His Gln Ile Ser Asp Val Ser	Glu Tyr Ile His Val				
	865	870			875	
Thr Arg Ser Leu	Leu Thr Ile Asp Cys Phe Ser	Phe Asp Ala Arg Ile				
	885	890			895	
Asp Thr Ala Thr	Tyr Ile Thr Val Asp Asp	Lys Tyr Leu Lys Val Ala				
	900	905			910	
Asp Phe Ser Leu	Gly Thr Glu Gly Ile Pro Ile	Pro Leu Val Asn Ser				
	915	920			925	
Val Asp Asn Asp	Glu Pro Pro Ser Leu Glu Tyr	Ser Lys Arg Arg Phe				
	930	935			940	
Gln Tyr Asn Asp	Gln Val Asp Ile Ser Ser	Val Ser Arg Asp Phe Cys				
	945	950			955	
Ser Gly Cys Ser	Cys Asp Gly Asp Cys Ser	Asp Ala Ser Lys Cys Glu				
	965	970			975	
Cys Gln Gln Leu	Ser Ile Glu Ala Met Lys Arg	Leu Pro His Asn Leu				
	980	985			990	
Gln Phe Asp Gly	His Asp Glu Leu Tyr Glu Ser	Ser Glu Lys Gln Asn				
	995	1000			1005	
Lys Phe Leu Lys	Leu Phe Phe Phe Arg Val Pro	His Tyr Gln Asn Arg				
	1010	1015			1020	
Leu Leu Ser Ser	Lys Val Ile Ser Gly Leu Tyr	Glu Cys Asn Asp Gln				
	1025	1030			1035	
Cys Ser Cys His	Arg Lys Ser Cys Tyr Asn Arg	Val Val Gln Asn Asn				
	1045	1050			1055	
Ile Lys Tyr Pro	Met His Val Ser Leu Phe Asn	Asp Asp Thr Tyr Gln				
	1060	1065			1070	
Leu Leu Phe Phe	Leu Gln Ile Phe Lys Thr Ala	Gln Ser Gly Trp Gly				
	1075	1080			1085	
Val Arg Ala Leu	Thr Asp Ile Pro Gln Ser Thr	Phe Ile Cys Thr Tyr				
	1090	1095			1100	
Val Gly Ala Ile	Leu Thr Asp Asp Leu Ala Asp	Glu Leu Arg Asn Ala				
	1105	1110			1115	
Asp Gln Tyr Phe	Ala Asp Leu Asp Leu Lys Asp	Thr Val Glu Leu Glu				
	1125	1130			1135	
Lys Gly Arg Glu	Asp His Glu Thr Asp Phe Gly	Tyr Gly Gly Asp Glu				
	1140	1145			1150	
Ser Asp Tyr Asp	Asp Glu Glu Gly Ser Asp Gly	Asp Ser Gly Asp Asp				
	1155	1160			1165	
Val Met Asn Lys	Met Val Lys Arg Gln Asp Ser	Ser Ser Glu Ser Gly Glu				
	1170	1175			1180	
Glu Thr Lys Arg	Leu Thr Arg Gln Lys Arg Lys	Gln Ser Lys Lys Ser				
	1185	1190			1195	
Gly Lys Gly Gly	Ser Val Glu Lys Asp Asp Thr	Thr Pro Arg Asp Ser				
	1205	1210			1215	
Met Glu Lys Asp	Asn Ile Glu Ser Lys Asp Glu	Pro Val Phe Asn Trp				
	1220	1225			1230	
Asp Lys Tyr Phe	Glu Pro Phe Pro Leu Tyr Val	Ile Asp Ala Lys Gln				
	1235	1240			1245	
Arg Gly Asn Leu	Gly Arg Phe Leu Asn His Ser	Cys Asp Pro Asn Val				

1250	1255	1260
His Val Gln His Val Met Tyr Asp Thr His Asp Leu Arg Leu Pro Trp		
1265	1270	1275
Val Ala Phe Phe Thr Arg Lys Tyr Val Lys Ala Gly Asp Glu Leu Thr		1280
	1285	1290
Trp Asp Tyr Gln Tyr Thr Gln Asp Gln Thr Ala Thr Thr Gln Leu Thr		1295
	1300	1305
Cys His Cys Gly Ala Glu Asn Cys Thr Gly Arg Leu Leu Lys Ser		1310
1315	1320	1325

<210> 28
 <211> 12700
 <212> DNA
 <213> Caenorhabditis elegans

<220>
 <221> CDS
 <222> (1001)...(1133)

<221> CDS
 <222> (4522)...(5208)

<221> CDS
 <222> (6128)...(6361)

<221> CDS
 <222> (7962)...(8350)

<221> CDS
 <222> (8706)...(8928)

<221> CDS
 <222> (9260)...(9516)

<221> CDS
 <222> (10328)...(10567)

<221> CDS
 <222> (11677)...(11700)

<400> 28
 aaaaatttttaa aaaaatttttt aaaaatttcgt gtaaaaaatta ccccggttgt ttaggaaata 60
 ataaagagat tagagacttt ttccagattt ttatttttctt gagttttgcc ggtttttcagc 120
 cgattttctat cttttttttt tcatttttttg tgattttttt tcgctagttt tcccctcaat 180
 ttctcgattt ttccacgatt ttctgaaaat ttccggaaaa ttgaattgtt tgcaaaaaaa 240
 aaaattcaaa aaccgcattt ttctcagaat ttctctggga ttttgtacaa atttttgaat 300
 tattttctcaa aaaaaagcag gtttttaccg atttttttgg ttttttcccc aaaattttcc 360
 gattttttcc gagttttgcc ggtttttcagc cgaattctac tctcgatttt ttacgattt 420
 ttgggaaatt ttccgaaaat tatttgaaaa aaaatcaaaa aaccgcattt ttttttctga 480
 attttctggg attttgtacg aaattttgaa atttttctcg aaaaaagcaa gttattcccc 540
 aaaattttct gattttcccc caaaaattta gatttttccc gagttttccc cagttctcag 600
 ctgattttcta tatttttttc tcaatttttg tgattttttg ttgctagttt tcccttcaat 660
 tctctgagtt ttccacgatt ttctggagat ttccgaaaaa ttggttgaaa aaaatcaaga 720
 aaccacattt ttctctggat ttctctgaaa ttgcacaaa atttttgaat tttttcgtaa 780
 aaaaaaactg ttttcccaa aaatttcaga ttgttttttg atttttttcg agattttccc 840
 ctgattttcaa agttttttcc tgaatttttc gaatttttcc tgaaaaatcg gctattttcta 900
 acttttttaa taattttttt tgaattttct acttttttaa tccttttttt tttgccattt 960

```

tttcccatct aaaattctaa attattcaaa attttacaga atg tca gaa gta atc 1015
Met Ser Glu Val Ile
1 5

gac gaa agt atc tta aat aca gaa gct tca gat gat cca ata cct cca 1063
Asp Glu Ser Ile Leu Asn Thr Glu Ala Ser Asp Asp Pro Ile Pro Pro
10 15 20

tta aat gat gat cag att gct gag ctt ttg ggt gaa gat gga gaa att 1111
Leu Asn Asp Asp Gln Ile Ala Glu Leu Leu Gly Glu Asp Gly Glu Ile
25 30 35

atg gag ata act gag cag aaa g gtgagatttt ttgagtaaaa ccttgaattt 1163
Met Glu Ile Thr Glu Gln Lys
40

tgcaactaaaa atttgcaatt ttcgctaaaa attaccttaa aactcgaaaa ttggaatttc 1223
tagctgagaa aatggccaaa aatgtcga aaacctgtga aaaaaaaac 1283
caccaaaaag gtttctaggc caccaaaaag atttctaggc caccaaaaat gtttctaggc 1343
caccaaaaat gtttctaggc caccaaaaat gtttctaggc caccaaaaat gtttctaggc 1403
caccaaaaat gtttctaggc caccaaacag gtttcaatgc caccaaaaat gtttctaggc 1463
caccaaaaat gtttctaggc ccccaaaaaa tttttctagg ccaccaaaaa gggttctagg 1523
ccaccaaaaa tggttctagg ccaccaaaaa gggttctagg ccaccaaaaa gggttcaatg 1583
ccaccaaaaa gggttctagg ccaccaacca gggttcaatg ccaccaaaaa tggttctagg 1643
ccaccaaaaa gggttctagg ccaccaaaaa tggttctagg ccaccaaaaa tggttctagg 1703
ccaccaaaaa gggttcaatg ccaccaaaaa gggttctagg ccaccaaaaa tggttctagg 1763
ccaccaaaaa tggttctagg ccaccaaaaa gggttctagg ccaccaaaaa gggttctagg 1823
ccaccaaaaa tggttctagg ccaccaaaaa gggttctagg ccaccaaaaa gggttcaatg 1883
ccaccaaaaa tggttctagg ccaccaaaaa gggttcaatg ccaccaaaaa tggttctagg 1943
ccaccaaaaa gggttcaatg ccaccaaaaa tggttctagg ccaccaaaaa gggttctagg 2003
ccaccaaaaa tggttctagg ccaccaaaaa tggttctagg ccaccaaaaa gggttctagg 2063
ccaccaaaaa gggttcaatg ccaccaaaaa tggttctagg ccaccaaaaa gggttcaatg 2123
ccaccaaaaa tggttctagg ccaccaaaaa tggttctagg ccccaaaaa atttttctag 2183
gccacaaaa aggttctag gccacaaaa atgttctag gccacaaaa aggttctag 2243
gccacaaaa aggttcaat gccacaaaa aggttctag gccacaaac aggttcaat 2303
gccacaaaa atgttctag gccacaaaa aggttctag gccacaaaa atgttctag 2363
gccacaaaa atgttctag gccacaaaa aggttctag gccacaaaa aggttcaag 2423
gccacaaaa aggttcaat gccacaaaa atgttctag gccacaaac aggttcaat 2483
gccacaaaa aggttctag gccacaaaa atgttctag gccacaaaa aggttctag 2543
gccacaaac aggttcaat gccacaaaa aggttctag gccacaaac aggttcaat 2603
gccacaaaa atgttctag gccacaaaa aggttctag gccacaaaa atgttctag 2663
gccacaaaa atgttctag gccacaaaa aggttctag gccacaaac aggttcaat 2723
gccacaaaa atgttctag gccacaaac aggttcaat gccccaaaa aatttttcta 2783
ggccacaaa aaggtttcta ggccacaaa aatgtttcta gaccacaaa aaggtttcta 2843
ggccacaaa aatgtttcta gaccacaaa aaggtttcta ggccacaaa aatgtttcta 2903
ggccacaaa aaggtttcta ggccacaaa aatgtttcta ggccacaaa aaggtttcta 2963
ggccacaaa caggtttcaa tgccacaaa aaggtttcta ggccacaaac caggtttcaa 3023
tgccacaaa aatgtttcta ggccacaaa aaggtttcta ggccacaaa aatgtttcta 3083
ggccacaaa aatgtttcta ggccacaaa aaggtttcta ggccacaaa aaggtttcaa 3143
ggccacaaa aaggtttcaa tgccacaaa aatgtttcta ggccacaaa caggtttcaa 3203
tgccacaaa aaggtttcta ggccacaaa caggtttcaa tgccacaaa aaggtttcta 3263
gaccacaaa aaggtttcta ggccacaaa caggtttcaa tgccacaaa aaggtttcta 3323
ggccacaaa caggtttcaa tgccacaaa aatgtttcta ggccacaaa aaggtttcta 3383
ggccacaaa aatgtttcta ggccacaaa aatgtttcta ggccacaaa aaggtttcta 3443
ggccacaaa caggtttcaa tgccacaaa aatgtttcta ggccacaaa caggtttcaa 3503
tgcccaaaa aaatttttct aggccaccaa aaaggtttct aggccaccaa aaatgtttct 3563
agaccaccaa aaaggtttct aggccaccaa aaatgtttct agaccaccaa aaaggtttct 3623
aggccaccaa aaatgtttct aggccaccaa aaaggtttct aggccaccaa acaggtttcta 3683

```

atgccaccaa	aaatgtttct	aggccaccaa	aaatgtttct	aggcccccaa	aaaatttttc	3743
taggccacca	aaaagggttc	aatgccacca	aaaatgtttc	taggccacca	aaaagggttc	3803
taggccacca	aaaatgtttc	taggccacca	aaaatgtttc	taggccacca	aaaagggttc	3863
taggccacca	aacagggttc	aatgccacca	aaaatgtttc	taggccacca	aacagggttc	3923
aatgccacca	aaaagggttc	taggccacca	aaaatgtttc	tagaccacca	aaaagggttc	3983
taggccacca	aacagggttc	aatgccacca	aaaagggttc	taggccacca	aacagggttc	4043
aatgccacca	aaaatgtttc	taggccacca	aaaagggttc	taggccacca	aaaatgtttc	4103
taggccacca	aaaatgtttc	taggccacca	aaaagggttc	taggccacca	aacagggttc	4163
aatgccacca	aaaatgtttc	taggccacca	aacagggttc	aatgccacca	aaaatgtttc	4223
taggccacca	aaaatgtttc	taggccccca	aaaaattttt	ctaggccacc	aaaaagggtt	4283
ctaggccacc	aaaaatgttt	ctagaccacc	aaaaagggtt	ctaggccacc	aaaaatgttt	4343
ctagaccacc	aaaaagggtt	ctaggccacc	aaaaatgttt	ctaggccacc	aaaaagggtt	4403
ctaggccacc	aaaaatgctt	ctaggccacc	aaaaatgttt	ctacgccacc	aaaagccgcc	4463
tcaagcccga	aaaatttgaa	tttcccgtct	aaaaaatcta	aaattttccg	atttttcag	4521
ac gaa tca	gat gat	gtg gtg	atg ctg	gac gac	gat gat	gac act
Asp Glu Ser	Asp Asp	Val Val	Met Leu	Asp Asp	Asp Asp	Asp Thr
45		50		55		60
ccg gaa ccg	att ctc	gtg att	gat atg	gat gag	gat gag	gat gtt act
Pro Glu Pro	Ile Leu	Val Ile	Asp Met	Asp Glu	Asp Glu	Asp Val Thr
	65		70			75
aca gat ggt	cct gaa	tct cag	gaa gag	ctg gct	gca gat	gct ccg gct
Thr Asp Gly	Pro Glu	Ser Gln	Glu Glu	Leu Ala	Ala Asp	Ala Pro Ala
	80		85		90	
cca gga gct	cca gaa	gct tca	gct cca	gct caa	gaa gcc	tca gaa gct
Pro Gly Ala	Pro Glu	Ala Ser	Ala Pro	Ala Gln	Glu Ala	Ser Glu Ala
	95		100		105	
tca gct ccg	gat caa	gaa gct	cca gaa	gtt cag	gat gtt	ccg gat tct
Ser Ala Pro	Asp Gln	Glu Ala	Pro Glu	Val Gln	Asp Val	Pro Asp Ser
	110		115		120	
tcg gga gct	cca gat	gct tca	gct cag	gct tca	gag gct	tct gat gct
Ser Gly Ala	Pro Asp	Ala Ser	Ala Gln	Ala Ser	Glu Ala	Ser Asp Ala
	125		130		135	140
tca gct cca	gaa gtt	cca gga	tct aca	gaa gct	cag gat	gct cag gat
Ser Ala Pro	Glu Val	Pro Gly	Ser Thr	Glu Ala	Gln Asp	Ala Gln Asp
	145		150		155	
gtt ccg gat	tct ttg	gga gct	tca gat	gct tca	gct caa	gaa att cca
Val Pro Asp	Ser Leu	Gly Ala	Ser Asp	Gly Ala	Ser Ala	Gln Glu Ile Pro
	160		165		170	
gaa gct cca	gaa gcc	cca gaa	gct cca	gaa atc	gcc gct	gaa atc gac
Glu Ala Pro	Glu Ala	Pro Glu	Ala Pro	Glu Ile	Ala Ala	Glu Ile Asp
	175		180		185	
gaa gaa gtg	ctg ctc	gcc gag	caa aat	gga gtt	ttg gac	gaa gga ttt
Glu Glu Val	Leu Leu	Ala Glu	Gln Asn	Gly Val	Leu Asp	Glu Gly Phe
	190		195		200	
gat gag act	gac gat	att atc	ata gaa	gaa gaa	gaa gct	gta gaa gaa gct
Asp Glu Thr	Asp Asp	Ile Ile	Ile Glu	Glu Glu	Glu Ala	Val Glu Glu Ala
	205		210		215	220

gaa gcc gtg gag cca cca att aac act gaa aat cag gaa aac gcg ctg 5096
 Glu Ala Val Glu Pro Pro Ile Asn Thr Glu Asn Gln Glu Asn Ala Leu
 225 230 235

gaa atg ctc gaa gag cgc ctc aag aag aat gaa gaa aag gaa att gtg 5144
 Glu Met Leu Glu Glu Arg Leu Lys Lys Asn Glu Glu Lys Glu Ile Val
 240 245 250

gag aaa agt gat gtg aag cca gag gat gaa gat att ata cat atg gag 5192
 Glu Lys Ser Asp Val Lys Pro Glu Asp Glu Asp Ile Ile His Met Glu
 255 260 265

acg gat tca gtt gaa a gtatgggctt ttttagctgg aaaacaggaa aaaagagcaa 5248
 Thr Asp Ser Val Glu
 270

aaaattgata catttccagc ttaaccaatc tttttttgag ttgtaaagcc tgaaaattga 5308
 gattttttgta ccaactttta tgataaagct gaaaaaaaaa ttaattttttt gacgaattttt 5368
 tagcggaac cctgaaaaca tgttttgtct gaaaaataca gaaaatcgtc actttttaca 5428
 ataaattcga gatttttagc tcaaaaatac aacattatag tgcaaaaatc tcagaaaaag 5488
 ccaaaaattt cattcaaaaa tctcaaaaaa agcagaaaatt ttactcaaaa tatctcagaa 5548
 aaagctaaaaa ttttcccaaa aaatcccaga aaaagcagaa ttttcattca aaattcccag 5608
 aaaaagctga taatttacta aacaatctca gaaaatgctg aaattttact caaaagtctt 5668
 cataaaaagc tgaaatttta ctttaaaagt ttaggaaatg ctgcaatttc acttaaaaat 5728
 cccaaaaaag ctaaaatttt cccaaaaaat ccagaaaaa gcagaaattt tactcgaata 5788
 tctcaaaaaa aaaaagctg aaatttcact caaaaatccc agaaaaagct aaaaatttac 5848
 taaaaaatct caaaaaaaa aacgctaaaaa tttcactcaa aaatctcaga aaaagctaaa 5908
 attttactcg aatatctcaa aaaaaaaaac tgaaattttc ctaaaaaatt tatgaaaaac 5968
 cgaaatttca cttaaaagtc tcataaaaag ccgaattttc ccaaaaaaat ccagaaaaa 6028
 gctaaaaatt tactttaaaa tctcatctgt aatttttagt taaaatctca gaaaaaccg 6088
 aaattttctt caaaaatttg ctgattttca aattttcag cg tca agc cgc aaa cgt 6144
 Thr Ser Ser Arg Lys Arg
 275

act ggc gga gcc aca agt ccg cgg agc ccg gct caa aaa cga cca aaa 6192
 Thr Gly Gly Ala Thr Ser Pro Arg Ser Pro Ala Gln Lys Arg Pro Lys
 280 285 290 295

cga cgt gtt caa acg tta tta aag atg cgt cag aat gca att gaa cta 6240
 Arg Arg Val Gln Thr Leu Leu Lys Met Arg Gln Asn Ala Ile Glu Leu
 300 305 310

ttg aca cga ctt tat ggc tca tgg gat gca caa ttg agc ctc tca aat 6288
 Leu Thr Arg Leu Tyr Gly Ser Trp Asp Ala Gln Leu Ser Leu Ser Asn
 315 320 325

ctt gag aca att cga ttg ttg ggt gtc aat aat aat agg aag ctt atc 6336
 Leu Glu Thr Ile Arg Leu Leu Gly Val Asn Asn Asn Arg Lys Leu Ile
 330 335 340

gaa att ttt gag gag aat gag caa g gttaaagcgt ttttaaatgc 6381
 Glu Ile Phe Glu Glu Asn Glu Gln
 345 350

tatgaaaact gacaaatttt cgataaaaaa acggattttt ggaagaaaat cgctgaaaa 6441
 ttcatgtttt tctgcaaatt ttgaccaaatt tcccaagaaa aatacgattt tttagtccga 6501
 aaatcctcca aaaagatttc taggccacca aaaaggtttc taggccacca agaaagtttc 6561
 taggccacca aagtatttat aggccaccta agatgtttct aggccacctg agatgtttct 6621

aggtcaccaa aaatgtttct cggtcaccaa aaatgtttca aggccaccga aaaggtttct 6681
 aggccaccta agtatttcta ggccaccta gatgtttcta ggccacctga gatgtttcta 6741
 ggtcaccaaa aatgtttcta ggttaccaaa aatgtttcaa ggccatcgaa aaggtttcta 6801
 ggccaccaaa gtatttctag gccacctaa atgtttctag gccacctgag atgtttctag 6861
 gtcacaaaa atgtttcaag gccaccgaaa aggtttctag gccacaaaa aggtttctag 6921
 gccacaaaa atatttctag gccacctaa atgtttctag gccacctgag atgtttctag 6981
 gccacctgag atgtttctag gccacctgag atgtttctag gtcacaaaa atgtttctcg 7041
 gtcacaaaa atgtttcaag gccaccgaaa aggtttctag gccacctaa tatttctagg 7101
 ccacctaaaga tgtttctagg ccacctgaga tgtttctagg tcacaaaaaa tgtttctagg 7161
 ttacaaaaaa tgtttcaagg ccatcgaaaa gggtttctagg ccacaaaagt attttctaggc 7221
 cacctaagat gtttctaggc cacctgagat gtttctaggc caccaaaaat gtttcaaggc 7281
 caccgaaaag gtttctaggc caccaaaaag gtttctaggc caccaaaaat attttctaggc 7341
 caccaaaaat gtttctaggc caccaaaaat gtttctaggc caccaaaaat gtatcaaggc 7401
 caccaaaaag gtttctaggc caccaaaaat gtttctaggc caccaaaaat gtttctaggc 7461
 caccaaaaat gtttctaggc caccaaaaag gtttctaggc caccaaaaag gtttctaggc 7521
 caccaaaaag gtttctaggc caccaaaaag gtttcaaggc caccaaaaag gtttctaggc 7581
 caccaaaaat gtttctaggc caccaaaaat gtttctaggc caccaaaagta ttttctaggcc 7641
 acctaaaagg tttctaggcc atcaaaaagg tttctaggcc atcaaaaagg atttctaggcc 7701
 accaaaaata tttctaggcc acctaaagatg tttctaggcc accagagtat ttctaggcca 7761
 cctaagagggt ttctgggcca tcaaaaagggt ttcaagtcca tcaaaaagggt ttctaggcca 7821
 ccaaaaagggt ttctaggcca ccgaaaagggt ttctaggcca ccaaaaagggt ttctagacca 7881
 cctaagacat ttctaggcca acaaaaagggt ttctaggcca ccaagaagcc gaaaaactgt 7941
 ctcaaattcg aattttgcag tg ctc aaa caa aaa gtg tcc gca ctg aca gaa 7993
 Val Leu Lys Gln Lys Val Ser Ala Leu Thr Glu
 355 360

gag ctg aaa aag gag aag ctg gct cac gcg gga acc cgt tca gca ttg 8041
 Glu Leu Lys Lys Glu Lys Leu Ala His Ala Gly Thr Arg Ser Ala Leu
 365 370 375

aaa gaa ttg act aat gaa ata act gga atg cgt gta caa atg aat aaa 8089
 Lys Glu Leu Thr Asn Glu Ile Thr Gly Met Arg Val Gln Met Asn Lys
 380 385 390

cta cgt tca atg gtc act cag cct acg act tcg aaa att att gat agt 8137
 Leu Arg Ser Met Val Thr Gln Pro Thr Thr Ser Lys Ile Ile Asp Ser
 395 400 405 410

ttt gtt caa cgt cat cag gct ttc gag cag caa caa caa ttc caa cac 8185
 Phe Val Gln Arg His Gln Ala Phe Glu Gln Gln Gln Phe Gln His
 415 420 425

caa cac cac caa cac cga cca ata atg ttg gct cca cgt cat cat ccg 8233
 Gln His His Gln His Arg Pro Ile Met Leu Ala Pro Arg His His Pro
 430 435 440

ccg ccg ccc ccg cat ttt aca ccg aat caa cgg gcg gcg gct ccg tat 8281
 Pro Pro Pro Pro His Phe Thr Pro Asn Gln Arg Ala Ala Ala Pro Tyr
 445 450 455

cat ccg aat atg gtt caa ccg aat cgt ctt gct gct atg cca cat aga 8329
 His Pro Asn Met Val Gln Pro Asn Arg Leu Ala Ala Met Pro His Arg
 460 465 470

aga ccg att att gga atg cag gtgaaaatgg aatgccatga aaatttcggg 8380
 Arg Pro Ile Ile Gly Met Gln
 475 480

-79-

Tyr

```

atcgatatttt tttgactgaa aaatgtctga aaaatcaaaa atttttagcta aaaattgaga 9626
atatttttgt ttaaaaaaaaa tcattgaaat tgatttttttt ttattccata aaaatctcgg 9686
aaaagtcaat tttcagtcatt aaatcttctg aaaatttatcc aaacaatggg attttctgaa 9746
atttttagctt aaaaattgag gatttcccggt ttttttcaga gaaattccat tacaatcgat 9806
ttttttactg aaaaatcctc tggaaattaa caaaaaccaa ataaaatgcc ctaatttttt 9866
tttaaatcca aaaattgttg gattttttca gaaaaaata ttttttcaat tgactggtgt 9926
ccaaaaaata tagaaaattc aaattttcca agaaaaatag caaaaaaaat gtaatttttg 9986
tctaacaaaa aaattgaata gcgcaaaatt aaattgtcgt tttttttaat ttccctccgg 10046
ttttgaaagg aaaaaattcc ataaaaatcg aaattttttg actgaaaaat ccatgaaaac 10106
tcgaattttg agtcaaaaat cctctgaaaa tgctccaaaa tatgagattt tctgaaattt 10166
catcaaaaat taagaatttc acggttttaa aaaaattcca ttaaaatcga tattttttcaa 10226
gtgaaaaatc tctggaaaac tcgatgtttg agtcaaaatt cgtctgaaaa tgctccttta 10286
aattgaaaaa ttgaaaaaaa aaccgcccac aatatttgca g aat atc caa gtg ttc 10342
                                     Asn Ile Gln Val Phe
                                     645

```

```

gtc caa gtg tca tct ctt aaa ttc act gga atg aac ggt tac ccg gat 10390
Val Gln Val Ser Ser Leu Lys Phe Thr Gly Met Asn Gly Tyr Pro Asp
                                     650                                     655                                     660

```

```

cca gaa gat cgt ata tca att gac tgg gga tgc tcg aaa ttg tgg cct 10438
Pro Glu Asp Arg Ile Ser Ile Asp Trp Gly Cys Ser Lys Leu Trp Pro
                                     665                                     670                                     675

```

```

tgt aag ccg aaa tct cat cac aaa ttc cgt gta cgc ttc cat caa gca 10486
Cys Lys Pro Lys Ser His His Lys Phe Arg Val Arg Phe His Gln Ala
                                     680                                     685                                     690

```

```

caa ctg ctg ccg aag aac gat cga att acg att gtg gct gtg gcg aag 10534
Gln Leu Leu Pro Lys Asn Asp Arg Ile Thr Ile Val Ala Val Ala Lys
                                     695                                     700                                     705                                     710

```

```

gat aaa act agc gga att att cac att tcg cag gtgaaaaatt ggaaaatttg 10587
Asp Lys Thr Ser Gly Ile Ile His Ile Ser Gln
                                     715                                     720

```

```

cacaaatcca gacaaaaaaa actgaaaaat cgaaaaaatt tttgtaattt tttgccgaaa 10647
acgaaaaatta aaaactgata aaaattgatt tttaaccgga aaatccctga aaaatcaaac 10707
atttttttgct aaaaattgag aattatacgg ttttggttaa aaaaaaacta tttaaaaaaa 10767
atattttttc tttaaaaatc tcaacaaaaa aaaaaccaat tttcattcag aaatccccc 10827
ggagaattgt caaaattttg ggaatactct gaaatttcga taaacacctc atttttgatt 10887
aaaattgatt ttttaactga aaaatccctt aaaaaacgaa tatttttagtt ttttcacaaa 10947
aaaatgtgca atttatctga aatttcagca aaaaaaatga aaaaaaaaaa ttccgaaatt 11007
aaaaactgat aaaaatcgat tttttacttg aaaaattcgt gaaaaatcaa acacattttt 11067
gctaaccatt gagaatatta cgattttgtg aaaaaaaaaa ccattaaaat tgatttttta 11127
ttcctaataa atgccagaaa aatcaatttt cagtcaaaaa tcaccggaaa attatcaaaa 11187
ttttgaggtt ttctgtgaaa tttcaagctg aaatttccat ttttgaataa aaaaaatgtg 11247
gctggattta aaaaaaaacc attaaaattg attttttaac tgaaaaatcc gtattttctt 11307
gaaatttcag gcaaaaaatg tcatttccga aattaaaaat tgcgacaaaa tcaataaaaa 11367
ttgatcaaat ttgcaaaaaa aaaaaaactt tcgcaaaaaa tccttaaaat ttacattttt 11427
caacaaaaac tcgaattttc agtcaaaaat tcgtctgaaa atgctccaaa atatgggatt 11487
ttttgaaatt ttagctaaaa attgagaatt gcacgggtatt tagagaggga aaaattccat 11547
aaaaatcgat attttcctct ttaaaatctc gaaaaaaatc atcaattttt attcaaaaat 11607
cccccccgga aaattgtcaa aatttttgaga tttttctgaa atttcacgca aaaattttca 11667
ttttttcag ccc acc ttc atc act ctc gaa tga tcgatctctt caggtcaaat 11720

```

Pro Thr Phe Ile Thr Leu Glu *
725

```
gcactttttt ctggattttt ttgttaaaaa atttgaaatt ctcgtgtttt ttcttctgaa 11780
aaattgcttt ttttgatttt ttctgtaatt ttttttttgt tgattttctt aattttttta 11840
attttcaaaa aatctttttc atctctttct ctctctctct gaatctcaat tttttcctga 11900
atttcccggt ttttttctga taattttcaa tttttctctg aatttttcta ttccccccgt 11960
tgtaatgcca aaatatgtgg taatttctcc ccattttttc gctttattac tattttattct 12020
attcaattgg tgctctctc aatgtgttgt atgaaaaaca ctgttttatg gaggttttgg 12080
agaattttga attttttcgt cgtgattttt attggttttc tttaccaatt caattttttt 12140
tttaattcga aaattttag aaattcactt ttgtagctta aaaaattaaa aattgagaaa 12200
atgtgttcaa aaatggcaaa gttttcgaaa ttttagtcta aaaaaagatt tttttaatat 12260
agaattttta aaaattagca cagaaaaatg ccgaaaaatt cgtaattttt catttaaaaa 12320
tgaaaaaaa aaaaaacaaa aaaaaaaaaa aaaaagaggg aaaaatccca ttaaaagtag 12380
ttttttgact gcaaaatcgt ctggaaatta acaaaattta aaaaaatctt ttttacagcc 12440
catcgtttcc aaaaaccaaa taaaatgcca aaaaaaaatt tttatgcaaa aattctggat 12500
ttttttccga ttttttcaaa aaattccccc ttctaaaaaa aatggtgaat ttgttcccaa 12560
aaacccaaaa tttgagattt tctaaaattt tggcaaaaat taagaatttc acggttttga 12620
gagggaaaaa ctccattaaa attgatgatt ttatgactaa aaattcctaa aaaatcaatt 12680
ttcagtcaaa aattaaattt 12700
```

<210> 29

<211> 728

<212> PRT

<213> *Caenorhabditis elegans*

<400> 29

```
Met Ser Glu Val Ile Asp Glu Ser Ile Leu Asn Thr Glu Ala Ser Asp
 1           5           10           15
Asp Pro Ile Pro Pro Leu Asn Asp Asp Gln Ile Ala Glu Leu Gly
 20           25           30
Glu Asp Gly Glu Ile Met Glu Ile Thr Glu Gln Lys Asp Glu Ser Asp
 35           40           45
Asp Val Val Met Leu Asp Asp Asp Asp Thr Pro Glu Pro Ile
 50           55           60
Leu Val Ile Asp Met Asp Glu Asp Glu Asp Val Thr Thr Asp Gly Pro
 65           70           75           80
Glu Ser Gln Glu Glu Leu Ala Ala Asp Ala Pro Ala Pro Gly Ala Pro
 85           90           95
Glu Ala Ser Ala Pro Ala Gln Glu Ala Ser Glu Ala Ser Ala Pro Asp
100           105           110
Gln Glu Ala Pro Glu Val Gln Asp Val Pro Asp Ser Ser Gly Ala Pro
115           120           125
Asp Ala Ser Ala Gln Ala Ser Glu Ala Ser Asp Ala Ser Ala Pro Glu
130           135           140
Val Pro Gly Ser Thr Glu Ala Gln Asp Ala Gln Asp Val Pro Asp Ser
145           150           155           160
Leu Gly Ala Ser Asp Ala Ser Ala Gln Glu Ile Pro Glu Ala Pro Glu
165           170           175
Ala Pro Glu Ala Pro Glu Ile Ala Ala Glu Ile Asp Glu Glu Val Leu
180           185           190
Leu Ala Glu Gln Asn Gly Val Leu Asp Glu Gly Phe Asp Glu Thr Asp
195           200           205
Asp Ile Ile Ile Glu Glu Glu Ala Val Glu Glu Ala Glu Ala Val Glu
210           215           220
Pro Pro Ile Asn Thr Glu Asn Gln Glu Asn Ala Leu Glu Met Leu Glu
225           230           235           240
Glu Arg Leu Lys Lys Asn Glu Glu Lys Glu Ile Val Glu Lys Ser Asp
```


- 82 -

705
Gln Pro Thr Phe Ile Thr Leu Glu
725

715

720

<210> 30
<211> 11
<212> DNA
<213> Caenorhabditis elegans

<400> 30
aagatatgtg t 11

<210> 31
<211> 11
<212> DNA
<213> Caenorhabditis elegans

<400> 31
aacttcaaaa t 11

<210> 32
<211> 11
<212> DNA
<213> Caenorhabditis elegans

<400> 32
cttataagtt t 11

<210> 33
<211> 11
<212> DNA
<213> Caenorhabditis elegans

<400> 33
ttttccaaaa a 11

<210> 34
<211> 11
<212> DNA
<213> Caenorhabditis elegans

<400> 34
ttttttaaga t 11

<210> 35
<211> 6403
<212> DNA
<213> Caenorhabditis elegans

<400> 35
aaggaattag actctttatc taaagtgaag aatgatcaat taagaagttt ttgtcccata 60
gaattaaata taaatggatc tcttggggca gaatctgatt tggcaacatt ttgcacttct 120
aaaactgatg ctgttttaat gacttctgat gatagtgtga ctggatcgga attatcccct 180
ttgggtcaaag catgcatgct ttcacaaat ggatttcaga atattagtag gtgcaaagaa 240
aaagacttgg atgatacctg catgctgcat aagaagtcag aaagcccatt tagagaaaca 300
gaacctctgg tgtcaccaca ccaagataaa ctcatgtcta tgccagttat gactgtggat 360
tattccaaaa cagtagttaa agaaccagtt gatcacgaggg tttcttgctg caaaaccaa 420

gattcagaca	tatactgtac	tttgaacgat	agcaaccctt	ctttgtgtaa	ctctgaagct	480
gaaaatattg	agccttcagt	tatgaagatt	tcttcaaata	gctttatgaa	tgtgcatttg	540
gaatcaaaac	cagttatatg	tgatagtaga	aatttgacag	atcactcaaa	atttgcattg	600
gaagaatata	agcagagcat	cggtagcact	agttcagctt	ctgttaatca	ttttgatgat	660
ttatatcaac	ctattgggag	ttcaggtatt	gcttcactct	ttcagagtct	tccaccagga	720
ataaaggtgg	acagtctaac	tctcttgaaa	tgcggagaga	acacatctcc	agttctggat	780
gcagtgctaa	agagtaaaaa	aagttcagag	tttttaaagc	atgcagggaa	agaaacaata	840
gtagaagtag	gtagtgcact	tcctgattca	ggaaagggat	ttgcttccag	ggagaacagg	900
cgtaataatg	ggttatctgg	gaaatgtttg	caagaggctc	aagaagaagg	gaattccata	960
ttgcctgaaa	gaagaggaag	accagaaatc	tcttttagatg	aaagaggaga	aggaggacat	1020
gtgcatactt	ctgatgactc	agaagttgta	ttttcttctt	gtgatttgaa	tttaaccatg	1080
gaagacagtg	atggtgtaac	ttatgcatta	aagtgtgaca	gtagtgggtca	tgccccagaa	1140
atttgtgtct	cagttcatga	agattattct	ggctcttctg	aaagttcaaa	tgatgaaagt	1200
gattcagaag	atacagattc	ggatgatagc	agtattccaa	gaaaccgtct	ccagtctggt	1260
gtggttgtgc	caaagaattc	tactttgccc	atggaagaaa	caagtccttg	ttcttctcgg	1320
agcagtcaaa	gttatagaca	ctattctgac	cattgggaag	atgagagatt	ggagtcaagg	1380
agacatttgt	atgaggaaaa	atttgaaagt	atagcaagta	aagcctgtcc	tcaaactgat	1440
aagtttttcc	ttcataaagg	aacagagaag	aatccggaaa	tttcttttac	acagtccagt	1500
agaaaacaaa	tagataaccg	cctgcctgaa	ctttctcatc	ctcagagtga	tggggttgat	1560
agtacaagtc	atacagatgt	gaaatctgac	cctctgggtc	acccaaattc	agaggaaacc	1620
gtgaaagcca	aaataccttc	taggcagcaa	gaagagctgc	caatttattc	ttctgatttt	1680
gaagatgtcc	caaataagtc	ttggcaacag	accactttcc	aaaacaggcc	agatagtaga	1740
ctgggaaaaa	cagaattgag	tttttcttcc	tcttgtgaga	taccacatgt	ggatggcttg	1800
cactcatcag	aagagctcag	aaacttaggt	tgggacttct	ctcaagaaaa	gccttctacc	1860
acgtatcagc	aacctgacag	tagctatgga	gcttgtgggtg	gacacaagta	tcagcaaaat	1920
gcagaacagt	atggtgggac	acgtgattac	tggcaaggca	atggttactg	ggatccaaga	1980
tcaggtagac	ctcctggaac	tggggttgtg	tatgatcgaa	ctcaaggaca	agtaccagat	2040
tccctaacag	atgatcgtga	agaagaggag	aattgggatc	aacaggatgg	atcccatttt	2100
tcagaccagt	ccgataaatt	tcttctatcc	cttcagaaaag	acaaggggtc	agtgcagca	2160
cctgaaataa	gcagcaattc	cattaaggac	actttagctg	tgaatgaaaa	gaaagatttt	2220
tcaaaaaact	tagaaaaaaa	tgatatcaaa	gatagaggc	ctcttaaaaa	aaggaggcag	2280
gaaatagaga	gtgattctga	aagtgatggg	gagcttcagg	acagaaagaa	agttagagtg	2340
agggtagagc	agggagagac	atcagtgcct	ccagggttcag	cactgggttg	gccctcctgt	2400
gtcatggatg	acttcaggga	cccacagcga	tgggaaggaat	gtgccaaagca	agggaaaaatg	2460
ccatgttact	ttgatcttat	tgaagaaaat	gtttatttaa	cagaaagaaa	gaagaataaa	2520
tctcatcgag	atattaagcg	aatgcagtgt	gagtgtacac	ctctttctaa	agatgaaaga	2580
gctcaagggtg	aaatagcatg	tggggaagat	tgtcttaatc	gtcttctcat	gattgaaagt	2640
tcttctcggt	gtccaaatgg	ggattattgt	tccaatagac	ggtttcagag	aaaacagcat	2700
gcagatgtgg	aagtcatact	cacagaaaag	aaaggctggg	gcttgagagc	tgccaaagac	2760
cttccttcga	acacctttgt	cctagaatat	tgtggagagg	tactcgatca	taaagagttt	2820
aaagctcgag	tgaaggagta	tgcacgaaac	aaaaacatcc	attactattt	catggccctg	2880
aagaatgatg	agataataga	tgccactcaa	aaaggaaatt	gctctcgttt	catgaatcac	2940
agctgtgaac	caaattgtga	aacccaaaaa	tggactgtga	acggacaact	gagggttggg	3000
ttttttacca	ccaaactggt	tccttcaggc	tcagagttaa	cgtttgacta	tcagttccag	3060
agatatggaa	aagaagccca	gaaatgtttc	tgcggatcag	ccaattgccg	gggttacctg	3120
ggaggagaaa	acagagtcag	catcagagca	gcaggaggga	aatgaagaa	ggaacgatct	3180
cgtaagaagg	attcagtgga	tggagagcta	gaagctctga	tggaaaatgg	tgagggtctc	3240
tctgataaaa	accaggtgct	cagcttatcc	cggctaattg	ttagaattga	aactttggag	3300
cagaaactta	cctgtctgga	actcatacag	aacacacact	cacagtcctg	cctgaagtcc	3360
tttctggaac	gtcatgggct	gtctttgttg	tggatctgga	tggcagagct	aggtgacggc	3420
cgggaaaagta	accagaagct	tcaggaagag	attataaaga	ctttggaaca	cttgcccatt	3480
cctactaaaa	atatgttgga	ggaaagcaaa	gtacttccaa	ttattcaacg	ctgggtctcag	3540
actaagactg	ctgtccctcc	gttgagtga	ggagatgggt	attctagtga	gaatacatcg	3600
cgtgtcata	caccactcaa	cacacctgat	ccttccacca	agctgagcac	agaagctgac	3660
acagacactc	ccaagaaact	aatgtttcgc	agactgaaaa	ttataagtga	aaatagcatg	3720
gacagtgcaa	tctctgatgc	aaccagtgag	ctagaaggca	aggatggcaa	agaggatctt	3780
gatcaattag	aaaatgtccc	tgtagaggaa	gaggagaaga	tgcagtcaca	acagctactc	3840
ccacaacagc	tgctgaatg	caaagttgat	agtgaacca	acatagaagc	tagtaagcta	3900

```

cctacatctg aaccagaagc tgacgctgaa atagagctca aagagagcaa cggcacaaaa 3960
ctagaagaac ctattaatga agaaacacca tcccaagatg aagaggaggg tgtgtctgat 4020
gtggagagtg aaaggagcca agaacagcca gataaaacag tggatataag tgatttgccc 4080
acaaaactcc tggacagttg gaaaagaccta aaggaggtat atcgaattcc aaagaaaagt 4140
caaaactgaaa agggaaaacac aacaactgaa cgaggaaggg atgctgttgg cttcagagat 4200
caaacacctg ccccggaagac tcctaatagg tcaagagaga gagaccaga caagcaaaact 4260
caaaaataaag agaaaaggaa acgaagaagc tccctctcac caccctcttc tgcctatgag 4320
cggggaacaa aaaggccaga tgacagatat gatacaccaa cttctaaaaa gaaagtacga 4380
attaaagacc gcaataaact ttctacagag gaacgccgga agttgtttga gcaagagggtg 4440
gctcaacggg aggctcagaa acaacagcaa cagatgcaga acctgggaat gacatcacca 4500
ctgccctatg actctcttgg ttataatgcc ccgcatcatc cctttgctgg ttaccacca 4560
ggttatccca tgcaggccta tgtggatccc agcaacccta atgctggaaa ggtgctcctg 4620
cccacaccca gcatggaccc agtgtgttct cctgtcctt atgatcatgc tcagcccttg 4680
gtgggacatt ctacgaacc cctttctgcc cctccaccag taccagtggg gccacatgtg 4740
gcagctcctg tggaaagttc cagttcccag tatgtggccc agagtgatgg tgtagtacac 4800
caagactcca gcgttgctgt cttgccagtg ccggcccccg gccagttca gggacagaat 4860
tatagtgttt gggattcaaa ccaacagtct gtcagtgtac agcagcagta ctctcctgca 4920
cagtctcaag caaccatata ttatcaagga cagacatgtc caacagtcta tgggtgtgaca 4980
tcaccttatt cacagacaac tccaccaatt gtacagagtt atgccagcc aagtcttcag 5040
tatatccagg ggcaacagat ttccacagct catccacaag gagtgggtgg acagccagcc 5100
gcagcagtga ctacaatagt tgcaccaggg cagcctcagc ccttgcaagg atctgaaatg 5160
gttgtgacaa ataactctctt ggatctgccc ccccccctc ctcccaaacc aaaaccatt 5220
gtcttacctc ccaactggaa gacagctcga gatccagaag ggaagattta ttactacat 5280
gtgatcacia ggcagactca gtgggatcct cctacttggg aaagcccagg agatgatgcc 5340
agccttgagc atgaagctga gatggacctg ggaactccaa catatgatga aaaccccatg 5400
aaggcctcga aaaagcccaa gacagcagaa gcagacacct ccagtgaact agcaaagaaa 5460
agcaaagaag tattcagaaa agagatgtcc cagttcatcg tccagtgcct gaacctttac 5520
cggaacactg actgcaaagt gggaagaatt accacaactg aagactttaa acatctggct 5580
cgcaagctga ctcacggtgt tatgaataag gagctgaagt actgtaagaa tcctgaggac 5640
ctggagtgca atgagaatgt gaaacacaaa accaaggagt acattaagaa gtacatgcag 5700
aagtttgggg ctgtttacaa acccaaagag gacactgaat tagagtgact gttgggccag 5760
gggtgggagga tgggtggtca ggtaagacag actctagggg gaggaatcc tgtggcctt 5820
tctgtccac cctgtcagc actgtgtctc tgatgatata tcaccctggg gaattcaacc 5880
ctgcagatgt caactgaagg ccacaaaaat gaactccatc tacaagtgat tacctagtgt 5940
tgagctgttg gcatgtggtt agaagccatc agaggtgcaa gggcttagaa aagaccctgg 6000
ccagacctga ctccactctt aaacctgggt cttctccttg gcggtgctgt cagcgcacag 6060
acccatgcgc atccccaccc acaaccttt accctgatga tctgtattat attttaattg 6120
atatgtgaat atattgaaaa taatttggtt tttcctgggt tttgtttggt tttcgttttg 6180
cttttagcct ctacatgcta ggatcacagg aagactttgt aaggacagtt taagttctcc 6240
tgcaagggtt aatttggtat catgtaaata ttccaaagca ggctgccttg tggttttggc 6300
cagccttggt ctatgttgat aagattgatt tactgtctaa aatcacttta ctttatccaa 6360
tttttactga actttttatg taaaaaata aaatcaatta aag 6403

```

<210> 36

<211> 1915

<212> PRT

<213> Caenorhabditis elegans

<400> 36

```

---Lys-Glu Leu Asp Ser Leu Ser Lys Val Lys Asn Asp Gln Leu Arg Ser
  1          5          10          15
Phe Cys Pro Ile Glu Leu Asn Ile Asn Gly Ser Pro Gly Ala Glu Ser
      20          25          30
Asp Leu Ala Thr Phe Cys Thr Ser Lys Thr Asp Ala Val Leu Met Thr
      35          40          45
Ser Asp Asp Ser Val Thr Gly Ser Glu Leu Ser Pro Leu Val Lys Ala

```

- 86 -

-87-

-87-

- 88 -

- 89 -

1905

1910

1915

